AUGUST
1960
VOLUME 24
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDWARD S. BORDIN, Editor
University of Michigan

BOYD MCCANDLESS, Associate Editor
University of Iowa

Advisory Editors: John W. Atkinson, *University of Michigan*. Frank X. Barron, *University of California, Berkeley*. John M. Butler, *University of Chicago*. Allen T. Dittmann, *National Institute of Mental Health*. Seymour Fisher, *Baylor University*. Harrison G. Gough, *University of California, Berkeley*. Robert B. Malmo, *McGill University*. Paul E. Meehl, *University of Minnesota*. Daniel R. Miller, *University of Michigan*. Ruth L. Munroe, *New York, N. Y.* Saul Rosenzweig, *Washington University*. Julian B. Rotter, *Ohio State University*. Nevitt Sanford, *University of California, Berkeley*. Seymour B. Sarason, *Yale University*. Robert R. Sears, *Stanford University*. Edwin S. Shneidman, *Veterans Administration, Los Angeles, California*. Read D. Tuddenham, *University of California, Berkeley*.

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Manuscripts submitted for *regular* or *early* publication, Brief Reports, and tests for review should be sent to the Editor, Edward S. Bordin, Department of Psychology, University of Michigan, Ann Arbor, Michigan. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors.

Style of Manuscripts. Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association (1957 Revision)*, obtainable for $1.00 from the Washington office of the APA.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Managing Editor.

American Documentation Institute. To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reprints. Authors receive fifty reprints gratis. Additional quantities may be ordered when returning proofs.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 576 pages. The subscription per year is $8.00, foreign $8.50, single number $1.50.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Second class postage paid at Lancaster, Pa.
© 1960 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

August 1960 Vol. 24, No. 4

Contents

ARTICLES

Dimensions of Social Desirability: *Samuel Messick* - - - - - - - - - - - - - 279
Some Concepts Concerning the Mental Health of the Individual: *Donald C. Klein* - - 288
Ego Strength and Conflict Discrimination: An Experimental Construct Validation of the Ego Strength Scale: *Maurice Korman* - - - - - - - - - - - - - - - - 294
The Duration of the Archimedes Spiral Afterimage in the Diagnosis of Brain Damage: *Jerome D. Schein* - 299
A Study of Identification in Male Homosexuals: *Judy Chang and Jack Block* - - - 307
The Aftereffect of Seen Movement and Brain Damage: *R. H. Day* - - - - - - - - 311
Psychotherapists' Clinical Judgments and Attitudes towards Patients: *Martin S. Wallach and Hans H. Strupp* - 316
Operant Conditioning of Hostile Verbs in Relation to Experimenter and Subject Characteristics: *Donald C. Ferguson and Arnold H. Buss* - - - - - - - - - - 324
The Effect of a Structured Inquiry on Rorschach Scores: *Melvin Zax and George Stricker* 328
Clinical Diagnosis by the IPAT Music Preference Test: *Raymond B. Cattell and Robert E. McMichael* - 333
Personality Correlates of Manifest Anxiety in Children: *Luciano L'Abate* - - - - - 342
A New Scale of Social Desirability Independent of Psychopathology: *Douglas P. Crowne and David Marlowe* - 349
Relationship of Family Disturbance to Cognitive Difficulties in a Learning-Problem Child: *Michael A. Wallach, David N. Ulrich, and Margaret B. Grunebaum* - - - 355
Case History Correlates of Three MMPI Profile Types: *Harold Gilberstadt and Jan Duker* - 361
The Effectiveness of the Bender-Gestalt in Differentiating a Flight Group from an Aggressive Group of Adolescents: *Loren V. Corotto and Robert H. Curnutt* - - - 368

BRIEF REPORTS

The EPPS Profile Stability Coefficient: *Donald W. Fiske, Kenneth Howard, and William Rechenberg* - 370
Manifest Needs and Manifest Anxiety: *Carl D. Williams, Robert B. Tallarico, and James T. Tedeschi* - 371
The Influence of Order of Administration on Self-Concept Measures: *David S. Palermo and John G. Martire* - 372
Wechsler-Bellevue Patterns of Psychopaths: *Leonard Kingsley* - - - - - - - - - 373
The Factorial Validity of Items on the IPAT Anxiety Scale: *A. W. Bendig* - - - - 374

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual.* Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No.... from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $— for microfilm or $— for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual.*

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 279–287

DIMENSIONS OF SOCIAL DESIRABILITY [1]

SAMUEL MESSICK
Educational Testing Service, Princeton, New Jersey

A high correlation has been repeatedly noted between the judged desirability of a personality characteristic denoted by an inventory item and the probability of the item's endorsement (cf. Edwards, 1957b; Wahler, 1958). Edwards (1953) has pointed out that, while this correlation suggests the possibility that traits judged to be desirable are prevalent or dominant in the group under study, it might also reflect a tendency to create a good impression or to respond, either deliberately or unintentionally, in a socially desirable manner. Instruments designed to measure this set to respond desirably were found to be substantially correlated with several clinical and personality scales (Edwards, 1954; Edwards, 1957b; Fordyce, 1956; Hanley, 1957) thus documenting the pervasive influence of social desirability on personality measurement. In view of this accumulating evidence, Edwards (1954) and others (cf. Heineman, 1953; Sisson, 1948) attempted to control for social desirability in constructing personality measures by pairing, in a forced-choice format, items with approximately equal desirability scale values.

This forced-choice method of control assumes that social desirability is a single unidimensional continuum, along which statements may be spaced and in terms of which items may be equated. If, however, judgments of social desirability were found to be multidimensional, then controls in terms of a single average scale might prove inadequate. Edwards (1953) obtained his social desirability scale values by the method of successive intervals (cf. Edwards, 1957a; Torgerson, 1958), a technique which combines individual ratings of desirability to produce a single scale describing, on the average, the pooled judgments of the particular group used. Since the subjects' (*Ss*') own opinions about the social desirability of traits were elicited in the rating task, one might expect various cultural and social background factors to produce differences in the scales obtained from different judging groups. However, there may also exist within a single group consistent individual differences with sufficient structure to warrant their consideration as different "points of view" with respect to desirability. The purpose of the present paper is to investigate the dimensionality of social desirability judgments and to explore the nature of any dimensions or "points of view" uncovered.

Edwards (1957b) and others (cf. Klett & Yaukey, 1959) have investigated the comparability of social desirability judgments obtained from different groups and have found generally high correlations among the various scales. On a college sample, for example, Edwards (1953) found no essential differences in desirability values as a function of age, sex, or education, and Klett (1957a), on a high school sample, concluded that there were no significant differences among socioeconomic groups, or between grades or sexes. The correlation between desirability scale values obtained from the college and high school samples was .94. Klett (1957b) also obtained desirability ratings from a sample of mental hospital patients: he found no essential differences in the desirability ratings

[1] This project was supported by the Office of Naval Research, under Contract Nonr-2214(00). The study was greatly facilitated by C. James Klett, who graciously supplied his raw data for reanalysis. The author also wishes to thank Ledyard Tucker for his helpful advice on the analysis and rotations, Henrietta Gallagher for supervising the computations, Henry Kaiser for arranging to have an oblimax rotation performed on the Illiac computer, and Douglas N. Jackson for his perceptive comments on the manuscript. Portions of this paper were read at the American Psychological Association Meeting in Cincinnati, 1959.

of psychotic and nonpsychotic subgroups, and scale values for his total hospital sample correlated .88 with the college values and .87 with the high school values. In evaluating this striking similarity in group opinions about the desirability of personality traits, it must be remembered that successive intervals scale values represent a kind of average rating for the group involved. Within-group ratings would need to be very homogeneous and between-group differences quite marked before a low correlation would be expected between such averages. Rosen (1956), for example, reported highly similar mean profiles for groups given the MMPI under self-appraisal and desirability instructions, but at the individual level the correlations between normal and desirability profiles were low and sometimes negative.

Despite high correlations between scale values obtained from different groups, evidence suggesting the inadequacy of a single global view of desirability has been noted. Klett (1957a) found that matched item pairs for the Edwards Personal Preference Schedule (EPPS) were less adequately matched in terms of the high school values. French (1958) noted that EPPS scores for basic airmen were significantly higher than published norms on traits presumably valued by the Air Force, suggesting that recruits were able to identify responses that get military approval and that the Air Force view of desirability may differ from that expressed in the college ratings. Klett (1957b), noting that some of the differences between hospital and college ratings were systematically related to scale content, found that EPPS mean scores for a sample of schizophrenics were significantly higher than college normative scores for needs judged to be more desirable by the hospital group. Klett (1957c) also divided the mental hospital sample into several dichotomous groups in terms of high and low scores on scales derived from the MMPI. For each breakdown, correlations were computed between median desirability ratings obtained from high and low scorers; and some of the coefficients, although still high (.68 and .71), were significantly smaller than those previously reported.

The present study carries the investigation of consistent differences in judging desirability back to the level of the individual ratings. The technique applied is essentially a factor analysis of intercorrelations among ratings of personality items. This approach, used previously by Morris and Jones (1955) in their study of value dimensions, yields factors which presumably reflect different points of view with respect to the attribute rated, in this case social desirability.

An alternative approach to this same problem would be to analyze "*Q* correlations" between people to isolate groups of similar raters and examine their median patterns. This use of *Q* correlations, however, depends upon the existence of relatively pure types of raters; whereas for the present approach, which seeks consistent judgmental tendencies, it is an empirical question whether or not such types exist, i.e., whether there is a simple structure among factor scores for the raters. In the present analysis dimensions of desirability can emerge even though raters are "complex," i.e., have high factor scores on more than one dimension, which is the usual case in the area of mental abilities. In this question of research strategy, it was considered more fruitful to deal with a structured sample of items than to depend upon the isolation of a typology of raters. In any event, factor spaces obtained from the two approaches are closely related, and the distribution of people in the space may be examined in the present method by the subsequent determination of factor scores.

Method

Subjects

The present study is a reanalysis of desirability judgments collected by Klett (1957b) from 118 manifestly disturbed mental hospital patients. For the present analysis 10 *Ss* were eliminated from the sample because of incomplete data.

Procedure and Analysis

The *Ss* had been asked to rate on a nine-category scale the desirability of 140 items purportedly representing 14 psychological needs. Categories ran from one for "extremely undesirable" through five for "neutral" to nine for "extremely desirable." These items had been previously scaled by Edwards (1954) in constructing measures for the EPPS.

In Klett's (1957b) analysis of this material, the scaling procedure of successive intervals (cf. Torger-

TABLE 1

PERSONALITY ITEMS SELECTED FOR FACTOR ANALYSIS

Item	EPPS Item No.	Scale	Median Rating
1	17A	Deference	7.13
2	178A	Endurance	8.28
3	114A	Dominance	5.97
4	2A	Deference	8.00
5	68A	Order	8.21
6	152A	Change	8.20
7	41A	Achievement	5.97
8	45A	Autonomy	4.10
9	14A	Exhibition	6.80
10	78A	Succorance	7.82
11	221A	Nurturance	6.74
12	5A	Autonomy	8.41
13	214A	Heterosexuality	7.97
14	11A	Achievement	7.56
15	222A	Change	6.31
16	98A	Succorance	6.30
17	39A	Exhibition	5.44
18	82A	Intraception	6.47
19	94A	Dominance	4.55
20	224A	Heterosexuality	5.60
21	168A	Endurance	7.07
22	42A	Deference	6.09
23	91A	Affiliation	7.49
24	169A	Heterosexuality	5.78
25	146A	Affiliation	6.81
26	33A	Order	7.32
27	51A	Achievement	8.40
28	190A	Aggression	5.24
29	102A	Intraception	7.64
30	180A	Aggression	6.32
31	70A	Autonomy	5.67
32	65B	Endurance	7.09
33	93A	Succorance	5.36
34	76A	Affiliation	8.24
35	49A	Exhibition	4.47
36	206A	Nurturance	8.08
37	185A	Aggression	3.28
38	18A	Order	7.21
39	166A	Nurturance	7.42
40	104A	Dominance	7.34
41	197A	Change	6.60
42	112A	Intraception	6.17

son, 1958) was used, which in effect simultaneously normalized the ratings of the 140 items on a common base line. The scale values of the items were essentially the means of these normalized ratings as measured on the common scale. In the present analysis, on the other hand, the original ratings of each item were intercorrelated and factor analyzed. In a sense the two analyses complement each other in that the first one dealt with means and variances to obtain an average scale for the group, and the second dealt with intercorrelations to obtain the structure of consistent individual differences in judgments.

To make the factor analysis of feasible size, 42 items were selected from the 140 available, three from each of the 14 needs represented. By examining the distribution of ratings for each statement, one item was chosen with a high median desirability value for each need, one with an intermediate value, and one with a low value. No attempt was made to select in terms of item content, except to require three items for each need. Since the needs differed in their judged desirability, the median values for items representing one need overlapped with those of another to cover a wide range of desirability. This variation in judged desirability within each need favored the appearance of a general desirability factor, but at the same time the choice of three items from each scale permitted the occurrence of clustering according to needs. EPPS item numbers and median desirability ratings are presented in Table 1 for the statements selected. Product-moment correlations among the ratings of these 42 items were factor analyzed by the grouping method (Thurstone, 1947), with communalities estimated by the largest correlation in each column.

RESULTS

After the extraction of nine factors, the residuals were found to be symmetrically distributed about zero with 93% of the coefficients included between .10 and −.10. Since further factor extraction would have generated small doublets of doubtful interpretation, only nine factors were retained for rotation. Although one of them (Factor IX) was quite large, accounting for approximately 42% of the common variance and 22.6% of the total variance, a general factor of desirability did not seem indicated. For comparison purposes, the first centroid factor was also extracted; it accounted for 22.8% of the total variance. The factors were rotated analytically to oblique simple structure using the oblimax criterion (Pinzka & Saunders, 1954), and some minor additional rotations were performed graphically. The final rotated loadings are given in Table 2. In the orthogonal solution obtained by the grouping method, the nine factors accounted for the following percentages of total variance, respectively: 4.03, 3.57, 3.39, 9.05, 3.19, 2.48, 2.57, 2.53, 22.65. These orthogonal factors were naturally modified, sometimes considerably, in the course of rotation, but they correspond roughly to the oblique factors in the order listed in Table 2. The intercorrelations among

TABLE 2

FACTOR LOADINGS ROTATED OBLIQUELY

Variable	Factor I	II	III	IV	V	VI	VII	VIII	IX
1	.22	.12	−.06	−.01	−.12	−.26	−.15	.35	.31
2	−.10	−.04	−.06	−.09	.21	.03	.01	.16	.39
3	.31	.06	.03	.19	.08	−.01	.03	.03	.10
4	.01	−.00	−.18	.11	−.06	.08	.19	.10	.40
5	−.23	−.00	−.00	.01	.02	.25	−.03	.14	.34
6	.05	−.00	.27	−.13	.03	−.18	.55	−.11	.51
7	.56	−.04	.18	−.09	.06	.07	−.01	−.05	.08
8	−.13	.03	.28	.04	.10	−.00	.08	.03	−.27
9	.04	−.04	.26	.04	.25	−.05	−.04	−.04	.20
10	−.04	.02	.14	−.11	.02	−.00	.01	.15	.31
11	.43	.04	.39	−.11	.04	.00	−.14	−.13	.03
12	−.19	−.04	.02	−.02	.09	.08	.18	−.01	.46
13	−.22	−.13	.19	.04	.11	.07	−.00	.08	.12
14	.23	−.17	.23	.02	.08	−.21	.10	.04	.36
15	.01	.06	.35	.07	−.03	.16	.18	.16	−.17
16	.10	.06	.67	.06	−.08	−.05	−.07	−.01	−.08
17	−.03	.25	.48	.07	.06	−.03	−.10	−.06	.12
18	.64	−.12	.04	.11	−.05	−.01	.11	.14	.02
19	.10	.04	.25	.22	.21	.02	−.13	−.02	−.19
20	−.03	.01	.06	.64	.00	.02	−.05	.05	−.11
21	−.21	.15	−.04	−.11	.13	.27	−.18	.10	.13
22	−.02	.24	−.09	.13	−.06	.05	−.01	.43	−.06
23	.03	.03	.14	.16	−.17	.31	.04	.03	.29
24	−.05	−.05	.10	.59	−.02	−.02	.10	−.04	−.01
25	−.00	−.01	−.00	−.11	−.01	.67	−.01	.03	−.11
26	−.01	−.02	−.02	−.12	.03	.02	−.01	.52	.02
27	−.12	−.01	−.02	−.04	−.03	.14	.10	.06	.56
28	−.20	.30	.13	.23	.07	.03	.05	−.16	.07
29	−.04	.08	.04	−.05	.12	.01	.25	−.05	.44
30	−.05	.09	−.01	−.04	.56	−.07	−.01	.05	.04
31	−.16	.06	−.07	.15	.44	.09	−.10	.01	−.08
32	.21	.23	−.22	.25	.00	−.06	−.16	−.14	.61
33	.06	.29	.51	.15	−.07	−.07	−.12	.01	−.01
34	−.05	−.05	.11	−.14	−.02	.20	.20	−.02	.48
35	.02	.58	.31	−.01	.05	−.05	.09	−.08	.07
36	.27	−.10	.11	.17	−.12	.03	−.11	.04	.45
37	−.16	.62	.08	−.03	.02	.11	.07	.07	−.10
38	.06	−.08	.08	.22	.10	.09	−.11	.35	−.03
39	−.13	−.04	.27	.08	−.01	.41	.07	−.13	.08
40	.05	−.14	−.00	−.02	.54	−.10	.10	−.10	.29
41	−.01	.02	−.04	.07	−.02	.10	.69	−.00	−.02
42	.03	.13	−.06	.01	.36	.19	−.04	−.00	.05

primary factors (Table 3) are generally positive and indicate the presence of a second-order structure, but enough coefficients are sufficiently low to suggest that a general factor of desirability does not emerge at the second-order level either.

These factors summarize certain consistencies among the original ratings and may be interpeted in terms of different viewpoints as to what is desirable. Factor loadings .25 or greater will be emphasized in the interpretations. In general the items do not cluster according to the specific scales they were written to represent, but it should be noted that there is no necessary relationship between these factors of *judged* desirability and the scales or item clusters determined from *responses* to the inventory. The fact that the items do not cluster according to specific needs when rated for desirability does not reflect upon the homogeneity of scales from Edwards' inventory.

TABLE 3

INTERCORRELATIONS AMONG PRIMARY FACTORS

Factor	I	II	III	IV	V	VI	VII	VIII	IX
I	1.00	.11	−.03	.11	.15	.19	.16	.17	.03
II	.11	1.00	−.05	.16	.10	−.18	−.03	−.21	−.36
III	−.03	−.05	1.00	.35	.22	.28	.08	.48	.30
IV	.11	.16	.35	1.00	.49	.20	.33	.24	.12
V	.15	.10	.22	.49	1.00	.29	.27	.24	.29
VI	.19	−.18	.28	.20	.29	1.00	.17	.49	.58
VII	.16	−.03	.08	.33	.27	.17	1.00	.30	.15
VIII	.17	−.21	.48	.24	.24	.49	.30	1.00	.48
IX	.03	−.36	.30	.12	.29	.58	.15	.48	1.00

Factor I

18. To like to put yourself in someone else's place and to imagine how you would feel in the same situation. (Int.) .64
7. To like to write a great novel or play. (Ach.) .56
11. To like to have your friends confide in you and tell you their troubles. (Nur.) .43
3. To like to settle arguments and disputes between other people. (Dom.) .31
36. To like to treat other people with kindness and sympathy. (Nur.) .27

The common behavioral element viewed as desirable on this particular dimension might be labeled Interpersonal Concern or Interpersonal Sensitivity. The determination of factor scores would presumably uncover individuals in the sample analyzed with consistent tendencies to rate these items highly desirable, even though the scale values for the total group are only moderate (see Table 1).

Factor II

37. To like to make fun of people who do things you regard as stupid. (Agg.) .62
35. To like to use words which other people often do not know the meaning of. (Exh.) .58
28. To like to criticize someone publicly when they deserve it. (Agg.) .30
33. To like to have your friends make a fuss over you when you are hurt or sick. (Suc.) .29
17. To like to be the center of attention in a group. (Exh.) .25

This dimension might be labeled Aggressive Arrogance or Criticalness. The presence of Items 33 and 17, however, suggests that arrogant and critical behavior might possibly be considered desirable in this instance because of the social attention thereby derived and the attendant gratification of certain narcissistic and exhibitionistic needs. An alternative and perhaps more descriptive factor label, then, might be Narcissistic Exhibitionism.

Factor III

16. To like to have friends who feel sorry for you when you are sick. (Suc.) .67
33. To like to have your friends make a fuss over you when you are hurt or sick. (Suc.) .51
17. To like to be the center of attention in a group. (Exh.) .48
11. To like to have your friends confide in you and tell you their troubles. (Nur.) .39
15. To like to participate in new fads and fashions. (Chg.) .35
35. To like to use words which other people often do not know the meaning of. (Exh.) .31

8, 6, 39, 9, and 19 also have loadings above .25.

This dimension, which appears to be somewhat related conceptually to Factor II as a different aspect of ego-supporting needs, might be labeled Emotional Dependence or perhaps Succorance. In this connection Item 15 is probably not viewed so much as an indication of the need for change as of the need for attention and support, so that in this case, as on some of the other factors, one need may be instrumental to the gratification of another. Thus the need for change may here be activated in the service of the need for succorance by *subsidiation* (Murray, 1938, p. 86), a process not considered in the development of scales for one need at a time.

Factor IV

20. To like to listen or to tell jokes in which sex plays a major part. (Het.) .64

24. To like to read books and plays in which sex plays a major part. (Het.) .59
32. To like to be able to put in long hours of work without distractions. (End.) .25

This factor appears to be a doublet representing Sexual Interests. The appearance of Item 32, and perhaps even the next three items in order (28, 19, 38), is interesting from a psychoanalytic viewpoint. The absence of Item 13 dealing with "love for the opposite sex" suggests a possible differentiation in the desirability of love relationships and substitute sex interests.

Factor V

30. To like to attack points of view that are contrary to yours. (Agg.) .56
40. To like to argue for your point of view when it is attacked by others. (Dom.) .54
31. To like to do things your own way without regard to what others may think. (Aut.) .44
42. To like to judge people by why they do something—not by what they actually do. (Int.) .36
9. To like to tell other people about adventures and strange things that have happened to you. (Exh.) .25

This factor might represent something like Intellectual Aggressiveness or Active Intellectual Independence.

Factor VI

25. To like to write letters to your friends. (Aff.) .67
39. To like to show a great deal of affection toward your friends. (Nur.) .41
23. To like to have strong attachments with your friends. (Aff.) .31
21. To like to stick at a job even when it may seem as if you are not getting anywhere. (End.) .27

1 and 5 also have loadings above .25.

This dimension, which might be labeled Interpersonal Involvement, reflects a viewpoint in which affiliative and affectionate behavior is judged desirable.

Factor VII

41. To like to move about the country and live in different places. (Chg.) .69
6. To like to travel and see the country. (Chg.) .55
29. To like to analyze your own motives and feelings. (Int.) .25

This factor appears to be a highly specific doublet reflecting Interest in Travel. The need for change may be involved in this dimension, but it is of interest to note that Item 15, dealing with participation "in new fads and fashions," received a low loading.

Factor VIII

26. To like to have your meals organized and a definite time set aside for eating. (Ord.) .52
22. To like to be in groups where someone else takes the lead in deciding what you are going to do. (Def.) .43
38. To like to have your life so arranged that things run smoothly and without any change in plans. (Ord.) .35
1. To like to conform to custom and avoid doing things that people you respect might consider unconventional. (Def.) .35

This dimension might be labeled Compulsive Conformity. It reflects a viewpoint in which conforming deference, a kind of detached disinterest and orderliness are considered desirable.

Factor IX

32. To like to be able to put in long hours of work without distractions. (End.) .61
27. To like to do your very best in whatever task you undertake. (Ach.) .56
6. To like to travel and see the country. (Chg.) .51
34. To like to be loyal to your friends. (Aff.) .48
12. To like to be able to come and go when you want to. (Aut.) .46
36. To like to treat other people with kindness and sympathy. (Nur.) .45
29. To like to analyze your own motives and feelings. (Int.) .44
4. To like to learn what great men have thought about various problems in which you are interested. (Def.) .40
2. To like to start in and keep working on some assignment until it is completed. (End.) .39
14. To like to be recognized as an authority in some job, profession, or field of specialization. (Ach.) .36

5, 1, 10, 40, 23, and 8 also have loadings above .25.

This large factor appears to reflect an Achievement-Oriented, Middle-Class Stereotype of desirable behavior or a kind of Protestant Ethic. This popular conception favors a hard-working (Items 32, 2), success-oriented (27, 14), self-broadening (6, 4), friendly (34, 36, 23, 10), personally independent (12 and perhaps 40), socially con-

ventional (1), orderly (5), responsible (8) orientation.

The Factors in Relation to the Group Desirability Scale

Median desirability ratings computed for the total sample are presented in Table 1 for each item. For correlational purposes, these median ratings are very similar to the more refined successive intervals scale values used by Klett (1957b). This median desirability scale was correlated with each of the nine factors in turn, and the coefficients were found to be −.04, −.65, −.30, −.33, −.11, .12, .19, .15, and .76, respectively. The multiple correlation for the nine factors in predicting the median values was .96; the β-weights were −.05, −.44, .15, −.03, .24, .06, .29, and .78, respectively. Thus, the vector of medians can be oriented in the nine-dimensional factor space in such a way as to account for almost all of its variance. This is not a necessary finding, since the 42 median values represent averages of 42 frequency distributions, and the means of these distributions were essentially eliminated in the present analysis of intercorrelations. However, one might expect the average desirability scale for the judging group to be related to the component dimensions, particularly to the larger or dominant factor, since a point of view prevalent in the sample under study should make a larger contribution to the average scale.

Response Biases and Possible Artifacts

In the analysis of ratings based on a fixed number of categories, there are several possible sources of bias which could produce artifacts in the factor space. For example, ratings of items with medians near either extreme of the scale must be skewed, and this differential skewness has a small but consistent effect on the corresponding correlation coefficients, which could generate spurious dimensions similar to the "difficulty" factors in mental test analysis. Such factors should be consistently related to extreme median values, as are Factors II and IX in the present analysis; the other seven factors, however, involve mostly items from the same upper-middle range (Categories 5 to 7). Factor II includes some, but not all, items near the undesirable extreme and correlates −.65 with the medians; Factor IX involves mostly desirable items and correlates .76 with the medians. However, if a rotation is made in the plane of Factors II and IX, essentially equating IX with the medians, the new Factor II would then include primarily Items 35, 37, and 32, which are fairly consistent psychologically with the old Factor II but which involve both extremes of the rating scale. This leaves Factor IX highly related to the medians, and possible spurious influences should be considered in interpreting it. However, it is doubtful that differential skewness would generate the largest factor, especially when skewness was not extreme in the present data: the rating distributions for the 42 items selected had no modes in either extreme category and only 14 modes in either Category 2 or 8.

Another source of potential bias stems from consistent rater preferences for specific categories, such as a predominant use of the extremes or the middle range in making judgments. Most of the raters in the present sample, however, appeared to use a wide range of categories, with only about a dozen *S*s showing noticeable preferences for specific categories or restricted ranges. Although no more than two of these *S*s appeared to have markedly similar response patterns, consistent but slight category biases operating throughout the sample could distort the factor space and should be considered in the interpretation. In this connection, it is interesting to not that, when Factors II and IX are rotated as mentioned above, the new Factor II has high loadings for items from both extremes and could reflect intensity of rating.

Such rating biases thus offer alternative interpretations for Factors II and IX, but since there is little evidence for their occurrence in the present data, the simple structure orientation is preferred.

Discussion

The finding of divergent viewpoints about social desirability is obviously highly dependent upon the sample of individuals studied. For example, one could probably assemble a group of people with sufficient homogeneity that their ratings of desirability would reflect only one consistent point of view. The issue

in question, however, is not whether a single scale can be obtained for some particular population, but whether or not such an average scale is of sufficient generality to use in personality measurement to control for the tendency to respond in a socially desirable way. The present finding of several different dimensions of desirability suggests that such an average scale would not provide an adequate control in individual assessment and would have inconsistent effectiveness from subject to subject. Equating in terms of a large dimension which seemed to reflect a popular set of values for Western culture, such as Factor IX in the present analysis, might prove adequate for studying interrelations among variables at the *group* level; but for *individual* assessment the controls could be quite ineffective, particularly for people whose perceptions of desirability differed markedly from the group consensus.

The multidimensionality of social desirability ratings, the major finding of the present study, naturally must be cross-validated on normal populations. The stability and generality of the particular nine factors found here would be of interest, but it might be expected that several new and different points of view would be uncovered by varying the type of subjects sampled and items rated. However, it must also be considered that wide group differences could occur with only minor attendant changes in a factor structure which spanned several points of view. Morris and Jones (1955), for example, found that, although the ratings of various value orientations differed considerably across cultures, the factors obtained proved to be very similar.

In a society with widely differing social and subcultural influences, it is not surprising that there should be several dimensions of social desirability. However, the present finding of as many as nine factors, on a fairly restricted sample of 42 items and 108 cases, suggests the possibility that the consistent variation is not so much among clearly defined social groups, such as high school students or basic airmen, as among individual points of view. It is true that this diversity in viewpoint might be related to peculiarities in the judging sample, but in that case the factors could be important in their own right in relation to psychopathology. However, as Edwards (1957b, p. 48) has noted, his instructions for rating social desirability are similar to Rosen's (1956) for personal desirability, in that the *S*'s *own* judgments of the desirability of traits in others are emphasized. It might be expected that individual differences in such personal views of desirability would generate many factors, even for relatively homogeneous, normal populations. Thus, since the term "social desirability" appears somewhat misleading, even when discussing Edwards' (1957b) results, it would seem appropriate to substitute another label (cf. De Soto, Kuethe, & Bosley, 1959; Rosen, 1956) which emphasizes differences in individual viewpoints of desirability.

The discussion thus far has emphasized the reduction of error variance in personality measurement by controlling the effects of the tendency to respond in a socially desirable way. This response tendency, however, may also be viewed as a stylistic personality variable (Jackson & Messick, 1958), and in this connection, rather than trying to eliminate the effect, attempts might instead be made to measure the individual's position on the different dimensions. If, for example, measures of differential tendencies to respond in various desirable ways can be formulated, they may be used on the one hand as personality variables in their own right and on the other as multiple control measures to partial differential desirability biases out of content scales. Since a reliable tendency to respond in desirable ways might in itself reflect certain traits or needs in the respondent, much research is needed to develop control techniques (cf. Helmstadter, 1957), which do not require, as partial correlation does, that the corrected content scale be uncorrelated with the response tendency. It must also be remembered that the factors discussed in the present paper represent dimensions of *judged* desirability, which may or may not be related to dimensions of the tendency to *respond* in desirable ways, if such exist.

The proposed treatment of desirability dimensions as personality variables is similar in many ways to the diagnostic use of "ideal-self-" descriptions, which have been widely employed in evaluating outcomes of psychotherapy (Rogers & Dymond, 1955). Osgood, Suci, and Tannenbaum (1957, Ch. 6) have

also utilized ratings on the Semantic Differential, which has a large evaluative component, to compare conceptual structures in different personalities and to study personality change during psychotherapy. Incidentally, this study of differences in individual conceptions appears fruitful even though comparisons at the group level show nearly identical factor structures for different cultures and pathologies (Osgood, Suci, & Tannenbaum, 1957, pp. 170–175). Thus, the presence of consistent individual differences in the perception of desirability offers the possibility that the seemingly objective task of rating the desirability of a trait may be useful in making inferences about the personality of the rater.

Summary

Attempted controls in personality inventories for the tendency to respond in a socially desirable way usually assume that the domain of desirability is unidimensional. The present study attempted to evaluate this assumption and to explore the nature of any multidimensional structure uncovered. Ratings of the desirability of personality items from the Edwards Personal Preference Schedule were intercorrelated and factor analyzed. Nine factors were extracted and rotated to oblique simple structure. The factors were labeled and interpreted in terms of different "points of view" as to what is desirable. The finding of several dimensions of desirability suggests that the equating of items in terms of a single average desirability scale would leave much of the consistent variance of the ratings uncontrolled. Such a control method would provide inconsistent effectiveness for individual assessment, depending upon the *S*'s conformity to the group consensus of desirability.

REFERENCES

De Soto, C. B., Kuethe, J. L., & Bosley, J. J. A redefinition of social desirability. *J. abnorm. soc. Psychol.*, 1959, **58**, 273–275.

Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

Edwards, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corp., 1954.

Edwards, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957. (a)

Edwards, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957. (b)

Fordyce, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

French, Elizabeth G. A note on the Edwards Personal Preference Schedule for use with basic airmen. *Educ. psychol. Measmt.*, 1958, **18**, 109–115.

Hanley, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, **21**, 391–397.

Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–454.

Helmstadter, G. C. Procedures for obtaining separate set and content components of a test score. *Psychometrika*, 1957, **22**, 381–393.

Jackson, D. N., & Messick, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

Klett, C. J. The stability of the social desirability scale values in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 183–185. (a)

Klett, C. J. The social desirability stereotype in a hospital population. *J. consult. Psychol.*, 1957, **21**, 419–421. (b)

Klett, C. J. The social desirability stereotype and some measures of psychopathology. *J. consult. Psychol.*, 1957, **21**, 450. (c)

Klett, C. J., & Yaukey, D. W. A cross-cultural comparison of judgments of social desirability. *J. soc. Psychol.*, 1959, **49**, 19–26.

Morris, C., & Jones, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523–535.

Murray, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana, Illinois: Univer. Illinois Press, 1957.

Pinzka, C., & Saunders, D. R. Analytic rotation to simple structure: II. Extension to an oblique solution. Res. Bull. 54–31. Princeton, N. J.: Educational Testing Service, 1954. (Multilith)

Rogers, C. R., & Dymond, R. F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1955.

Rosen, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151–158.

Sisson, E. D. Forced-choice—the new Army rating. *Personnel Psychol.*, 1948, **1**, 365–381.

Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Wahler, H. J. Social desirability and self-ratings of intakes, patients in treatment, and controls. *J. consult. Psychol.*, 1958, **22**, 357–363.

(Received August 12, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 288–293

SOME CONCEPTS CONCERNING THE MENTAL HEALTH OF THE INDIVIDUAL [1]

DONALD C. KLEIN
Human Relations Service of Wellesley, Massachusetts

The concept of mental health defies definition despite efforts to render it operational and objective. The intent of this paper is to suggest an alternative set of concepts, which may prove more useful to the growing field of research and technology committed to mental health pursuits. The need for such concepts is clear. The urge of man to mold and perfect life has in the past two decades shifted, particularly in Western lands, from an attack on sociopolitical and economic forces to the intrapsychic and interpersonal forces believed to prevent the attainment of "the good life" (Duhl, undated). An example of this is given in the following quotation from a brochure issued by a mental health association in one of the northeastern states: "IF EVERYONE ENJOYED MENTAL HEALTH, there would be no crime, no insanity, and no deepseated unhappiness born of a feeling of frustration or of a sense that one does not belong." It would appear that we have by now badly overloaded the concept of mental health to the point where it has become virtually useless either for action or research.

Much of the early epidemiological work in this field has been concerned with cases of demonstrable pathology, i.e., patients accepted for hospital or outpatient care. From the viewpoint of such studies, mental health is the absence of mental illness, but even here there is the problem of determining generally acceptable criteria for deciding what is or is not an instance of genuine pathology. Problems multiply when we turn our attention from hospitals and clinics to the general population. The few statistics regarding distribution of problems in the general population seriously challenge our present inadequate concepts of mental health and emotional illness. For example, the "psychopathology of everyday life" has been demonstrated recently by Leighton (1956), who found that psychiatric appraisal of a sample of the population of a Canadian town revealed psychiatric impairment among at least 37% of the citizens.

Jahoda (1955) and Lindemann (1953), among others, have laid down explicit criteria for positive mental health. Their definitions describe the mental health of the individual in a social context. They seek to avoid enclosing him in the strait jacket of statistical normality or requiring him to acquiesce to malign environmental circumstances. In their definitions, adequate perception of reality, ability to adjust to inevitabilities and to master resolvable difficulties, and—for Lindemann—attainment of personal satisfactions and tranquility without impairment of others' functioning are viewed as prime characteristics. Thus, mental health is defined not as the absence of symptoms, but rather as the presence of three qualities in sufficient abundance: good reality testing, active mastery of the environment, and a feeling of contentment not based on victimization of others. At this point we are left with many questions regarding the interrelations of the three qualities. Does a deficiency in one imply that an individual no longer enjoys good mental health? Or must all three be reduced for ill health to ensue? Is any one—reality testing, for instance—a prerequisite for the other?

Such questions arise, in part, because three different basic meanings have been attached to the concept of mental health. In order not to perpetuate such confusion, the balance of this paper will concern itself with

[1] While maintaining responsibility for the assertions contained herein, the author wishes to express appreciation to Erich Lindemann, whose perspective and great vision have led to this attempt at clarification. Gratitude is also given to Frank Babbott for assistance with epidemiological concepts, to Frank and Joan Babbott for editorial help, and for their critical comments and suggestions to L. Borstelmann and W. Hollister.

the three connotations of mental health, using a different term for each. First, there is that meaning attached to the term "mental health" which appears to refer to the more enduring condition of the personality (hereafter referred to as "soundness"); second, there is that meaning which appears to refer to the more immediate state of health, equilibrium, or current effectiveness of the person in terms of the reaction of the organism to a stressor agent (hereafter referred to as "well-being"); third, there is that meaning which appears to refer to the ability of the individual to avoid illness even when exposed to illness producing agents (hereafter referred to as "emotional stability").

Soundness

By "soundness" is meant the level of integration of the general, more enduring personality structure. It is assessed in the perspective of the culturally shared ideal of the optimal in growth for any age level or other significant role dimension. Thus, in our culture it might include such attributes as adaptability, well developed social skills, a high degree of initiative, a sense of social responsibility, and a general perfusion of modest optimism in one's perception of man's struggle with nature and himself.[2] It is believed that such ideal attributes, though culture bound, tend to vary inversely with severity of psychiatric pathology. Perhaps psychiatric nosology is itself culture bound. Possibly such ideal characteristics are indeed hallmarks of emotional maturity, reflecting the superiority of our culture over others. It seems more probable, however, that in this society—as in most, if not all, others—ideal types have emerged which exemplify one or more of the *several sound integrations* of which man is biologically capable as he attempts to satisfy needs, pattern behavior, form social groupings, and find secondary satisfactions in interpersonal contacts.

Certain possible universal *attributes* of integration suggest themselves. The writer believes the attributes of an individual with a high soundness level can be expressed in his own terms pretty much as follows: "I feel I belong; I am a worthwhile person; I know where I stand." The external observer might choose to express the first attribute in terms of the clarity of the social role of the individual and the value ascribed to that role by the individual and others. He might then speak of the personal sense of worth of the individual and add an assessment of the appropriateness of role behavior as compared with the expectancies of the significant reference group.[3]

Well-Being

The concept "well-being" refers to the current state of equilibrium between the individual and the social-emotional environment impinging upon him. It is postulated that almost all individuals exhibit virtually continuous fluctuations in well-being. It is further believed that nearly all individuals undergo repeated states of ill-being and that such states are manifested in one or more of the commonly recognized patterns described in articles on combat fatigue, disaster fatigue, or psychological shock reactions observed in air raids and other mass stress situations (e.g., Janis, 1951). Finally, it has been noted in work with individuals undergoing emotional crises that states of ill-being tend to be time limited, the tendency being to return to a state of equilibrium whether or not treatment is available. The time period appears to depend upon the nature, severity, and duration of the stress.

Those who cannot arrive at an equilibrium following a state of ill-being are apt to suffer a serious decrement in soundness. This observation is supported by experience with

[2] A somewhat similar conception is presented by Shoben (1957) in a model of integrative adjustment. He also includes self-control which, as will be seen, this writer believes is more appropriately considered separately as one characteristic of stability.

[3] Some reinforcement for such speculations is offered by Schutz (1958), who suggests, on the basis of a large number of studies of group behaviors, that three core tasks in almost any interpersonal context relate to the questions: (a) Do I have or can I gain membership? (inclusion); (b) What is the structure or pattern of relationship into which I must fit? (control); and (c) To what extent can I safely express my personal needs in this relationship? (affection). Schutz further presents a schema in which each of the three interpersonal need areas is seen to vary according to whether it is handled in an "ideal" fashion, is "deficient" or "excessive" to a moderate degree, or is dealt with pathologically. Thus a pathological underdevelopment is believed to accompany schizophrenia.

combat fatigue cases during World War II, who developed a higher proportion of chronic emotional problems when removed permanently from combat units as compared with similar cases which were quickly returned to their units after a brief respite during the Korean conflict. Therefore a mental health program concerned with prevention and health maintenance is justified in devoting itself to the brief care of those suffering temporary impairment of well-being in an effort to forestall reductions in soundness levels.

When is a condition of ill-being in a *population* serious enough to warrant attention? This is one of the most challenging questions facing the field of community mental health. It can be satisfactorily answered only when base or "endemic" rates of prevalence and incidence for specific states of ill-being have been established among major population groupings. Meanwhile, many of those working with more or less closed communities, such as mental hospitals, schools, and prisons, are usually very clear in their own minds when institutional tensions affecting the well-being of large numbers in their care have reached "epidemic" proportions.

In order to establish base rates, it is necessary for there to be something to count. In one study, which led ultimately to a clearly defined preventive program, it was possible to count the number of withdrawals of students from nursing training. Generally acceptable indices of emotional distress are not always so readily available. And yet common sense and current societal preoccupations suggest that they do exist. Thus, for example, acts of violence, such as murder or assault, should justifiably be counted as instances of poor mental "health" (i.e., emotional ill-being) whether or not they occur in individuals of high or low soundness. Rates of occurrence of such behavior may also be taken as indications of the general mental "health" (well-being) of a population. Such indices may be of greater import in the evaluation of mental health programs than figures concerning cases treated in and out of the mental hospital. So it is important to study acute states of ill-being or crises *on their own merits* (irrespective of degree of more central personality dysfunction) in order to learn something about their etiology in the individual and their frequency and effects in the population.

Emotional Stability

The distinction between general soundness and well-being, the latter viewed as response to environmental stress, is derived from the ecologic notion that health or illness are states of equilibria between an organism and its environment (Burnet, 1953). The ability of the individual to cope with stress determines whether or not well-being is affected. While emotional well-being can be maintained under either ordinary or especially stressful circumstances, ill-being can result only from untoward reaction to stress. *Stability thus becomes the ability of the individual to cope with specific environmental stresses while maintaining a state of emotional well-being.* Stability levels of individuals can be compared inferentially by noting their behaviors under similar stressor conditions. Ultimately it may be possible to estimate stability levels by psychological laboratory tests, which have been found to correlate highly with adequacy of response to the specific stress or class of stresses in question.

As Whitehorn (1956) emphasizes, individuals differ markedly from one another in respect to situations they *perceive* as stressful. Given such variability, a phenomenological approach to the concept of stress suggests itself. Nevertheless, it is believed preferable for present purposes to consider stress in terms of forces impinging upon the individual, without regard to his reaction. Such a differentiation of stress from the organism's reaction permits clear-cut identification and ultimate categorization of psychological stressor agents. In preventive work the presence or absence of the stress in a population is worthy of study, apart from the reaction in any particular case.

Borrowing from the field of epidemiology, we have attempted to distinguish between four types of emotional stability.[4]

General Stability

The "general stability level" refers to the capacity of the individual to maintain equi-

[4] In epidemiology it is usual to think of immunity in four categories: inherited, general immunity; acquired, general immunity; inherited, specific immunity; and acquired, specific immunity.

librium or well-being under conditions of stress without regard to the nature of the stress itself. A "specific stability area," on the other hand, refers to the ability to withstand a specific kind of emotional stress. Any specific area of stability, therefore, must be defined in terms of some specific stressor agent. Specific stability is discussed more fully below.

Much work needs to be done in order to arrive at an adequate appraisal of the ingredients of the general stability level. In this regard, it should be worthwhile to examine the possible general "immunizing" effects of certain kinds of prolonged, insight-oriented psychotherapies. It may ultimately be demonstrated, for example, that classic psychoanalysis represents not only a therapy, aimed at relieving a specific state of emotional ill-being, or psychological exercise, which raises the general soundness level, but also a form of "immunization," which helps the individual ward off a wide variety of potentially noxious stimuli by increasing general stability or supporting specific stability areas.

Evidence has been advanced (Eysenck, 1952), which suggests that recovery or improvement rates are roughly 60% regardless either of the nature, depth, or length of the treatment or of whether it is carried out by psychoanalysts or family physicians (who may prescribe a few weeks in the general hospital for rest and recuperation). Such comparisons obviously suffer from the fact that uniform criteria of successful treatment do not exist. Nevertheless, once such criteria are available, it is not outside the realm of probability that so-called insight therapies will be found to resemble many other palliatives in the degree to which they restore emotional well-being, but will be proved significantly better able to render patients more generally stable under subsequent stress conditions, with the result that such patients are less apt to suffer recurring states of emotional ill-being.

The distinction between general stability and soundness is fuzzy at best. It is believed that soundness has a moderately high positive correlation with general stability but that the relationship is not univocal. It is assumed that high general stability can exist in an individual of only moderate soundness, especially when the personality defenses of the individual are such as to insulate him from the full impact of stress. Such defenses may not always be considered to reflect a high soundness level though they may be very effective aids to general stability. They include such reactions as repression, denial, selective inattention, and possibly even an ability under some circumstances to resort to the use of phantasy when reality becomes too bleak. Examples of apparently effective use of both denial and phantasy have been observed in some adult polio patients in respirators. These individuals appear to avoid the full impact of their predicament, whereas other patients not capable of so doing display depression and other signs of being overwhelmed by the all-too-real catastrophe.

Specific Stability

Studies of populations undergoing marked shifts in sociocultural environments may shed increasing light on the relationship between general and specific stability levels. It seems probable that individuals with high general stability will demonstrate areas of low specific stability when exposed to certain forms of stress. Careful review of soldier and civilian populations in wartime, of forced migrations of whole peoples, and of groups known to be either upwardly or downwardly mobile from a socioeconomic standpoint may suggest fruitful lines of future inquiry.

In order to study areas of specific stability, it is necessary ultimately to arrive at some categorization of types of stress conditions to which large segments of the population are routinely exposed. Some suggestions are to be found in the literature. Birth trauma is given some emphasis by certain psychoanalytic theorists. Perhaps here is the biologic prototype of the entry of the individual into a new, unfamiliar, confusing, and potentially harmful setting. Reactions to frustration or interference with goal-directed behavior have also been studied in the laboratory and find their counterparts in clinical patients, who become anxious or depressed when advancement up the economic or social ladder seems no longer possible. Lindemann's (1944) observations of reactions to bereavement have led to a formulation of emotional hazard which emphasizes sudden alterations or disruptions in key role relationships.

Inherited Stability

The distinction between inherited and acquired immunity—however useful for public health—may seem simply another way to raise the hackneyed controversy of nature vs. nurture, heredity vs. environment. It may be possible, nonetheless, to make some meaningful distinctions without promoting further useless controversy. At the outset, it may be helpful to remind ourselves that behavior does not exist in the absence either of the organism or its environment.

There are those (e.g., Erikson, 1950) who suggest that children are born with temperamental response tendencies (heredity, if you will) which affect parental responses to them (i.e., environment) in various ways, depending in part upon what the mother herself brings to the relationship. If the mother's initial attempts at mothering are discouraged by certain response tendencies in the child, permanent damage may be done to the affectional ties between them. A recent study by Klebanoff (1959) supports this viewpoint. He found that mothers of children with brain damage and of schizophrenic children responded similarly to questions on a child rearing attitude scale and that both, in turn, showed significantly more pathological attitudes toward child rearing and the family than did mothers of normal children.

If some children are born with response tendencies rendering them highly vulnerable to maternal rejection, they may be described as having inherited a kind of emotional instability and therefore as being susceptible to subsequent emotional maladjustment. Impairment of the mother-child relation as a result of the way in which the mother handles the child's inadequacy will result in a child with an extremely low soundness level (Fries, 1944; Mittelman, 1954). Obviously such speculations regarding inherited stability and instability must be tested ultimately by the detailed observations of newborn children, such as those of Fries (1944) who correlated response tendencies of neonatal infants with the specific mothering patterns required by each type of child to insure healthy psychological development.

Acquired Stability

In epidemiology the concept of acquired immunity implies that the organism has had some experience with the stressor agent and, in the process, has developed successful (i.e., nonillness producing) ways of coping with it for some period of future time. During the period of immunity, the organism and the agent are able to coexist without undue harm to the former, or the invader is destroyed.

In this area, as in so many instances where we are concerned with the relationship between sickness and health, it is desirable to begin first with the observations of groups of people who have become sick under demonstrable stress conditions. Having noted the reactions of those who become markedly ill, it should become possible to shift our attention to those whose emotional well-being is only slightly or temporarily impaired by similar stress. Finally, it may be possible to discover situations in which small amounts of stress render individuals increasingly able to cope with more massive doses of similar stresses in the future. The following statement by Frank (1957) from a national magazine expresses the belief that sibling rivalry, known in its virulent form to play a part in disturbing emotional equilibrium, may under certain conditions act so as to confer a kind of stability in a specific area:

> In families where each child feels he has reasonable access to his parents' attention and affection, sibling rivalry is not only inevitable but desirable. It is inevitable because every child, at least initially, wants to be his parents' favorite. It is desirable because every child is bound to meet conflict sooner or later, at school if not before, and there is no better place in which to learn how to cope with it constructively than in the home where there is love and concern for both parties. Sibling rivalry is one of the major ways by which children can develop a sense that the world is, within limits, a trustworthy place in which they can feel reasonably secure (p. 59).

This point of view must be examined and carefully tested within the framework of specific responses to specific stress conditions. Otherwise, there is the possibility that it will gain popular support prematurely and join the growing list of changing fashions in child rearing. The danger of such uncritical acceptance is readily seen after consideration of the probability that, in some circumstances, small stresses do not immunize, but rather cumulate and eventually overwhelm the individual.

In the search for specific areas of acquired

stability, initial priority must go to the search for stressor agents, a search that is well underway. Ultimately, however, it becomes necessary to relate major stress areas, which have been shown to produce casualties, to prototypical situations in the early life development of the organism that occasion mild stress without disabling the child. For example, while all of us learn in a variety of ways a multiplicity of social skills, relatively little has appeared in the literature regarding such skills and how they are established.[5] In this search, it may be desirable to examine a wide variety of disturbance of parent-child relationships noted in the child guidance literature because each may have its counterparts in the repertoire of role interactions of most, if not all, parents and children. Such a search might have two desirable results: first, it might suggest conditions under which stressor agents exceed a safe tolerance level; second, it might point to conditions under which individuals are not exposed to sufficient amounts of stress, thus rendering them less stable in similar situations in later life when more enduring personality patterns have been developed.

Summary

This discussion has put forth the suggestion that the search for definitions of mental health be abandoned, at least temporarily, in favor of a differentiated approach to the problem of human adjustments and maladjustments. It has been suggested that current concepts of mental health and illness are primarily concerned with the long standing emotional or psychological condition of the individual, for which the term "soundness" is proposed. The immediate state of "well-" or "ill-being" of the individual at any particular time is proposed as suitable for more intensive study, both because of its possible effect on the soundness level in some instances, and because it should be considered in and of itself as a period of acute illness or malfunctioning and worthy of note in population studies. The vulnerability of the individual to environmental stress has been discussed in terms of a possible "general stability factor," related to but not identical with soundness level, and of areas of "specific stability" which offer a promising field for further study in terms of possible implications for child rearing and educational practices. Perhaps by following this more differentiated approach to mental health, we may find that apparently conflicting doctrines of today may be viewed as much less contradictory in a more adequate viewpoint of tomorrow.

[5] One notable exception to this statement is Cameron's (1947) work on the social origins of mental illnesses, which suggests that emotional breakdown is essentially founded upon inadequate or inappropriate social perceptions and skills.

References

Burnet, F. M. *The natural history of infectious disease.* (2nd ed.) Cambridge: Cambridge Univer. Press, 1953.

Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.

Duhl, L. Psychiatry and the community. Unpublished Manuscript, National Institute of Mental Health, undated. (Mimeo.)

Erikson, E. H. *Childhood and society.* New York: Norton, 1950.

Eysenck, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.

Frank, J. D. Are you a guilty parent? *Harper's*, 1957(April), **214**, 56–59.

Fries, M. Psychosomatic relationships between mother and infant. *Psychosom. Med.*, 1944, **6**, 159–162.

Jahoda, Marie. Toward a social psychology of mental health. In A. M. Rose (Ed.), *Mental health and mental disorder.* New York: Norton, 1955.

Janis, I. *Air war and emotional stress.* New York: McGraw-Hill, 1951.

Klebanoff, L. B. Parental attitudes of mothers of schizophrenic, brain injured and retarded, and normal children. *Amer. J. Orthopsychiat.*, 1959, **29**.

Leighton, D. C. The distribution of psychiatric symptoms in a small town. *Amer. J. Psychiat.*, 1956, **112**, 716–723.

Lindemann, E. Symptomatology and management of acute grief. *Amer. J. Psychiat.*, 1944, **101**, 141–148.

Lindemann, E. Mental health—fundamental to a dynamic epidemiology of health. In I. Galdston (Ed.), *The epidemiology of health.* New York Acad. Med., 1953.

Mittelman, B. Motility in infants, children, and adults. *The psychoanalytic study of the child.*, 1954, **9**, 142–177.

Schutz, W. *FIRO: A three-dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

Shoben, E. Toward a concept of normal personality. *Amer. Psychologist*, 1957, **12**, 183–189.

Whitehorn, J. C. Stress and emotional health. *Amer. J. Psychiat.*, 1956, **112**, 773–781.

(Received July 7, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 294–298

EGO STRENGTH AND CONFLICT DISCRIMINATION: AN EXPERIMENTAL CONSTRUCT VALIDATION OF THE EGO STRENGTH SCALE [1]

MAURICE KORMAN
University of Texas Southwestern Medical School

Barron's *Es* scale was validated against the predictive criterion of successful outcome in psychotherapy (Barron, 1953). Because of the method of its construction, its content and correlates, it is suggested that it measures ego strength, a presumably latent trait often obscured by the vividness of psychopathology. Research with the *Es* scale has tested its criterion validity by examining its predictive and postdictive power (Wirt, 1955, 1956) and investigated its construct validity principally by computing its correlation with indices based on other psychometric instruments such as the Rorschach and the Bender-Gestalt (Tamkin, 1957; Tamkin & Klett, 1957). This type of "psychometric" construct validation presents a number of difficulties. First, such tests frequently can marshal but little empirical or theoretical evidence for being acceptable operational specifications of the construct in question, a drawback not remedied by a plethora of low but positive correlations (see Jenkins & Lykken, 1957). Secondly, as Campbell and Tyler (1957) have pointed out, intertest correlations may represent common but construct-irrelevant features (e.g., response sets). They add: "Construct validation is epitomized by the correlation between two or more specifications of a construct maximally *different* in apparatus or method" (p. 91). In Cronbach's terms (1957), the present study attempts to incorporate two basically divergent modes of approach, the correlational and the experimental.

[1] This research was carried out at the Psychopathic Hospital of the State University of Iowa. Carl Bickert's help in running subjects is gratefully acknowledged.

A version of this paper was given at the Annual Meeting of the APA, Cincinnati, 1959.

Because the *Es* scale is empirically defined it is especially desirable for other specifications of ego strength to be linked to a promising theoretical network. A congruence between such measures would then go beyond indicating that certain face valid experimental procedures relate to the construct interpretation of the test; it would serve as a bridge relating the psychometric device to an existing theoretical network. Dollard and Miller's explication of ego strength in terms of higher mental processes (1950) was taken as the starting point. The components of higher mental processes are many; principally, they include the ability to appropriately mediate discrimination, generalization, transfer, learned drives, and reinforcement. In this study the capacity of an individual to resolve a discrimination conflict was chosen as the experimental index of ego strength. The reasons for this choice were twofold: first, there is an extensive literature beginning with Pavlov (1927) on the conflictful characteristics of difficult discriminations; secondly, since the increased ability to solve conflicts and to make appropriate discriminations are both among the commonly accepted goals of psychotherapy, a more felicitous parallel between the experimental procedures and the *Es* scale is achieved. In addition, the parameters of discrimination problems are sensitive to experimental manipulation; one can readily increase the conflict loading of a discrimination task by placing the subject (*S*) under instruction to be as fast as he can while at the same time informing him of the irrevocable nature of his decision once it is made.

The hypothesis was therefore formulated that a high *Es* group should be able to resolve

discrimination conflict more speedily than a low *Es* group. The differences between the *Es* groups should be greatest for the condition of extreme discrimination difficulty (the most conflict provoking) and decrease as the discrimination becomes easier. In the case of easiest discrimination some general disruption of performance might still be expected because of the generalization on conflict (Brown, 1942), although the prediction here is less clear.

Procedures

A discrimination situation involving the method of constant stimuli was used, patterned after Festinger (1943) and Block and Petersen (1955). The *Ss* sat in a darkened room approximately 4½ feet away from a black box, behind which were two lights placed about 10 inches apart and shining through holes 9/32 of an inch. The top light (standard stimulus) was a 7½-watt opaque bulb, while the variable light was a regular 40-watt white bulb. The voltage applied through the variable light at 7 different rheostat positions was 60, 70, 78, 80, 82, 90, and 100 volts, respectively. In addition the voltages were chosen in such a way as to make the two lights nearly equal in intensity when 80 volts were applied at the variable light. The S sat behind a table with his preferred hand on a toggle. He was instructed to make a judgment as to whether the variable light was brighter or dimmer than the standard light and to indicate his choice by pushing the toggle in the appropriate direction. In order to increase the difficulty of the task, a .5-sec. interval occurred between the standard and the variable light, at the end of which a clock was activated which ran until the *S* pushed the toggle to the left or to the right. He was further told: "I'd like you to make up your mind as soon as it's possible for *you* but keep in mind that once you've pushed the lever up or down you can't change your mind again."

Each of the seven intensities of the variable stimulus were presented a total of 15 times by using three successive random orderings of the stimuli and by presenting this complete set of 21 trials five times.

Subjects

Forty-seven recently admitted psychiatric inpatients were used as *Ss*. Excluded were patients with a diagnosis or suspicion of organic brain disease, patients undergoing some type of radical therapeutic interference such as electroshock or insulin, and patients too uncooperative, agitated, or withdrawn to participate. The *Ss* (31 women and 16 men) were administered the MMPI and their *Es* scores obtained through the Hathaway and Briggs norms (1957). The entire range of *Es* scores rather than the extremes was used in this study so as to make it a more severe test of the scale's power. Because of the sex-linked speed elements of the task, the male and female *Ss* were considered separately. A cutting score (*T* score) of 41, the median, was used to separate high and low female *Es* groups. Since the *Es* scores for the smaller sample of males appeared to be markedly skewed, the same cutting score was adopted for them.[2]

Results

Prior to any analysis of the data, rules were laid down for the identification of the "easy," "intermediate," and "difficult" conflict situations based on *Ss*' individual performances. The criterion used involved each *S*'s consistency in labeling any one of the seven intensities of the variable stimulus brighter (or dimmer) than the standard stimulus. Those intensities of the variable stimulus which were most consistently seen as brighter (or dimmer) and which were furthest away from the point of objective equality (the middle intensity) were chosen as representing the least difficult discrimination for that *S*. Likewise, the most difficult discriminations were those which were most inconsistently seen as brighter (or dimmer) and which were nearest to the point of objective equality. The third category, that of intermediate difficulty, was defined by choosing those intensity positions which fell between the easy and difficult discriminations with reference to the criteria outlined above. Each of these three difficulty levels was represented by two intensities, 15 judgments being given to each. The mean of the 30 prejudgment latencies at each difficulty level was the basic data of this study. That no distortion of the data occurred through these procedures is shown by the fact that extreme positions were chosen as easy discrimination cells while the center positions tended to become difficult discrimination cells; the high and low *Es* groups did not differ in this respect.

As can be seen from Table 1 the high *Es* group holds a few more psychotics and is slightly older than the low *Es* group; these differences are far from significant (for diagnosis, chi square = .18 with 2 *df*; for age, $t = .83$ with 43 *df*).

[2] Due to circumstances beyond control, fewer subjects than desired were run. The statistical analyses were performed on females and males separately, the latter group serving essentially as a small replication sample.

TABLE 1

CHARACTERISTICS OF HIGH AND LOW *Es* GROUPS

Variable	High *Es*	Low *Es*
Diagnosis[a]		
Psychotics	40%	33%
Neurotics	28%	33%
Personality Disorders	32%	33%
Age		
Mean	31.85	28.32
SD	13.90	13.39
Es		
Mean	52.25	31.10
SD	7.75	5.76

[a] Based on $N = 43$; diagnosis on four patients was deferred.

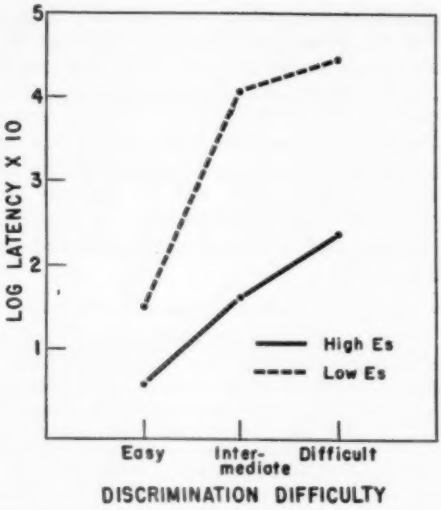

FIG. 1. Mean prejudgment latencies of *Es* groups for increasing levels of discrimination difficulty.

The mean latency scores of the high and low *Es* groups on easy, intermediate, and difficult discriminations are shown in Fig. 1. The low *Es* group has a longer latency throughout, although the difference becomes more marked at the intermediate and difficult levels. Figure 1 combines male and female data because of the remarkable similarity of their separate curves within each *Es* group.

In order to test the significance of the differences between the curves of Fig. 1, exclusive of the easy discrimination category, the log transformed latency scores were subjected to separate analyses of variance for males and females after it had been concluded, by means of the Bartlett test, that the assumption of homogeneity of variance was not being violated. In order to compensate for the correlation introduced by repeated measurements on the same *Ss*, the analysis suggested for such cases by Edwards (1951) was used. Table 2 reports these results: in the case of both males and females there is a significant main effect between *Es* groups ($p = <.03$ for the females; $p = <.06$ for the males). The low ego strength patient takes longer in resolving the discrimination conflict than does the high ego

TABLE 2

ANALYSES OF VARIANCE FOR REPEATED MEASURES OF LOG TRANSFORMED LATENCY SCORES FOR HIGH AND LOW *Es* GROUPS

Source	Females			Males		
	df	Mean Square	*F*	*df*	Mean Square	*F*
Between groups	1	.5618	4.94**	1	.3298	4.34*
Between *Ss* same Group	29	.1138		14	.0760	
Between Trials	2	.0674	5.14**	2	.0279	
Trials × Groups	2	.0075		2	.0022	
Pooled *Ss* × Trials	58	.0131		28	.0281	

* $p = .06$.
** $p = .03$.

TABLE 3

RESULTS OF *T* TESTS BETWEEN HIGH AND LOW *Es* GROUPS AT THREE LEVELS OF DISCRIMINATION DIFFICULTY

Discrimination Difficulty	*t*	
	Females (29 *df*)	Males (14 *df*)
Easy	1.05	1.78*
Intermediate	2.13**	1.87*
Difficult	2.10**	2.07*

* $p < .10$.
** $p < .05$.

strength patient. The differences between the two levels of discrimination difficulty were significant at the .03 level for the females but did not differ significantly for the males.

As Fig. 1 suggests, there is a disproportionate increase in latency for the low *Es* group as the discrimination difficulty moves from easy to intermediate and difficult as compared to the nearly linear relationship exhibited by the high *Es* group; the differences between easy and intermediate discriminations are significantly different for the two groups ($p < .02$) as are the differences between easy and difficult discriminations ($p < .03$). The groups do not differ, however, in their increment from intermediate to difficult discriminations. These findings generally received support from a comparison of the *Es* groups' performance on the three levels of discrimination difficulty. Table 3 reports the obtained *t*'s for males and females separately. In the case of the females, while the high and low *Es* groups do not differ on the easy discrimination, they differ significantly on both the intermediate and difficult levels of discrimination difficulty ($p < .05$). This trend is obscured somewhat in the case of the males because of the smaller *N;* the male high and low *Es* groups differ at the .10 level only with reference to all three levels of discrimination difficulty.

DISCUSSION

This study corroborates the findings of Tamkin (1957) and Tamkin and Klett (1957) with regard to the *Es* scale's inability to distinguish degrees of psychopathology as discriminated by psychiatric diagnosis. Contrary, however, to Tamkin and Klett's view that this is a "serious challenge to its validity," the present writer would suggest that this lack of association with overt psychopathology is a predictable and necessary by-product of the *Es* scale's construction which ignored the initial psychiatric status of the patients in favor of their treatment outcome. The difficulty, of course, lies in the surplus meanings of the term "ego strength" which lead to paradoxical conclusions.

The hypotheses concerning the relationship of the *Es* scale to discrimination conflict were generally borne out. In two separate samples, patients with low *Es* scores (a measure of ego strength, psychometrically defined) took longer than high *Es* patients to resolve a discrimination conflict (a measure of ego strength, experimentally defined). The latency between stimulus and response had been chosen as the dependent variable in this study because of its sensitivity to conflict instructions. The question arises as to whether the low *Es* group also made more errors in the successive discrimination judgments than the high *Es* group. Comparing the two groups on their total number of errors for the six intensities of the variable stimulus (leaving out the objectively "equal" center set) yields a nonsignificant *t* of .14 with 45 *df*. Since there is no a priori reason for suspecting personality variables to be related to the *judgments* in a psychophysical exercise, the null hypothesis can reasonably be accepted.

The *Es* scale's empirical derivation in the context of psychotherapeutic outcome had previously suggested a tentative identification of its underlying construct with ego strength, independently of any consideration of ego strength as one of the higher mental processes in the Dollard and Miller sense. The results of this study suggest that the "theory" underlying the *Es* scale is one which appropriately includes the construct ego strength as it is defined in the Dollard and Miller framework. Reciprocally, it is also possible to look upon these data as lending validity to an explication of ego strength patterned after Dollard and Miller and couched in terms of appropriateness of discrimination and conflict resolution. This can be maintained because part of the variance in such behaviors can be ac-

counted for by the variance of *Es* scores which are in turn associated with personality characteristics that make for a better "therapy bet."

With reference to the postulated relationship between *Es* scores and levels of discrimination difficulty, the results are likewise in a predicted direction although not in all respects. As the discrimination becomes more difficult, the low *Es S*'s performance becomes disproportionately worse, while the high *Es S*'s latency increases in a more stable, linear fashion. However, as Fig. 1 shows, the low *Es S*'s latency on the most difficult discriminations is not very much greater than his latency for discriminations of intermediate difficulty. A number of possibilities present themselves. Low *Es* Ss possibly make only the distinction between quite easy and all other discriminations; this would suggest another type of inappropriate generalization on their part. It is also not improbable that their latency on discriminations of intermediate difficulties is close to the asymptote they can reach under the particular instructions given, i.e., to be both speedy and correct, and that a further increase in the difficulty of the discrimination has little effect.

Summary

As a test of the construct validity of Barron's *Es* scale, 47 psychiatric inpatients were placed in a discrimination situation based on the method of constant stimuli. The hypotheses were (*a*) that a high *Es* group should be able to resolve a discrimination conflict more speedily than a low *Es* group and (*b*) that group differences should increase as the difficulty of the discrimination becomes greater.

The hypotheses were generally supported. The construct which the *Es* scale measures is related to ego strength as defined in the Dollard and Miller framework in terms of capacity for appropriate discrimination and conflict resolution.

REFERENCES

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

Block, J., & Petersen, P. Some personality correlates of confidence, caution and speed in a decision situation. *J. abnorm. soc. Psychol.*, 1955, **51**, 34–41.

Brown, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, **31**, 272–292.

Campbell, D. T., & Tyler, B. B. The construct validity of work group morale measures. *J. appl. Psychol.*, 1957, **41**, 91–92.

Cronbach, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671–684.

Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

Edwards, A. E. *Experimental design in psychological research*. Rinehart: New York, 1951.

Festinger, L. Studies in decision: I. Decision time, relative frequency of judgements and subjective confidence as related to physical stimulus differences. *J. exp. Psychol.*, 1943, **32**, 291–306.

Hathaway, S. R., & Briggs, P. F. Some normative data on new MMPI scales. *J. clin. Psychol.*, 1957, **13**, 364–368.

Jenkins, J. J., & Lykken, D. T. Individual differences. In P. R. Farnsworth (Ed.), *Annual Review of Psychology*. Vol. 8. Palo Alto: Annu. Reviews, 1957. Pp. 79–112.

Pavlov, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.

Tamkin, A. S. An evaluation of the construct validity of Barron's ego-strength scale. *J. clin. Psychol.*, 1957, **13**, 156–158.

Tamkin, A. S., & Klett, C. J. Barron's ego strength scale: A replication of an evaluation of its construct validity. *J. consult. Psychol.*, 1957, **21**, 412.

Wirt, R. D. Further validation of the ego strength scale. *J. consult. Psychol.*, 1955, **19**, 444.

Wirt, R. D. Actuarial prediction. *J. consult. Psychol.*, 1956, **20**, 123–124.

(Received July 16, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 299–306

THE DURATION OF THE ARCHIMEDES SPIRAL AFTERIMAGE IN THE DIAGNOSIS OF BRAIN DAMAGE [1]

JEROME D. SCHEIN
Florida State University

Since Price and Deabler (1955) introduced the Archimedes spiral afterimage as a technique for differentiating brain damaged patients from other neuropsychiatric patients, there has been an increasing accumulation of evidence suggesting that the diagnostic efficiency of this test is far below that which clinicians would find useful as a routine screening device (Berger, Everson, Rutledge, & Koskoff, 1958; Gilberstadt, Schein, & Rosen, 1958; Goldberg & Smith, 1958; Philbrick, 1959). However, these investigators have relied on a pass-fail criterion for each trial, requiring only that the subject report whether or not he sees the negative afterimage (henceforth referred to as AIR for "afterimage report"). On the other hand, attempts to explain the failure of brain injured patients to report this phenomenon have implied that damage to the cortex reduces the *duration* of the afterimage (henceforth referred to as AID for "afterimage duration") with complete failure to report the afterimage representing an endpoint on the duration continuum.

Three theories have recently been proposed to account for the tendency of brain damaged individuals to report the negative afterimage induced by visual stimulation less frequently than patients with intact cerebrums. Saucer (1953) and Saucer and Deabler (1956) consider the perception of the afterimage to be a function of a global effort of the entire cortex. Applying a principle of mass action, they conclude that brain injury would foreshorten AID. Klein and Krech (1952) use the Pavlovian principle of inhibition to support their position that brain injury reduces AID, though their own research involved tactual aftereffects. Shapiro (1954), dealing directly with the perception of apparent movement, also applies the concept of inhibition in predicting that brain injury will foreshorten AID.

Though the structure of their theories differ somewhat, these theorists are unanimous in predicting the effect of brain damage on AID. They provide the basis for avoiding the embarrassment of an all-or-none explanation of the earlier studies; i.e., one can account for brain damaged individuals who do give an AIR by assuming that they have a reduced AID, albeit not reduced below their perceptual limen. Thus, because duration is reduced, brain injured individuals should give fewer AIRs.

The empirical evidence on spiral AID is mixed. Eysenck, Holland, and Trouton (1957) asserted that some depressant drugs have effects isomorphic with brain injury, in that both increase inhibitory potentials in the brain. When they administered a placebo, a stimulant, and a depressant drug to 6 normal subjects (*Ss*), they found that the depressant drug (sodium amylobarbitone) significantly reduced AID, though the stimulant did not increase it. They felt that this result supported those theories of brain functioning which make use of the concept of cortical inhibition.

Three studies of brain injured patients

[1] This research is a portion of a PhD thesis submitted to the Department of Psychology, University of Minnesota. The author is indebted to William Schofield, Ephraim Rosen, Paul E. Meehl, and David T. Lykken of his dissertation committee for guidance and criticism. Appreciation is also due the staff of the Minneapolis VA Hospital for their unfailing cooperation and encouragement.

have included some observations on AID. Gallese (1956) noted that the false negatives (brain damaged patients who did report the afterimage) seemed to experience a less marked and apparently shorter AID than control patients. Page, Rakita, Kaplan, and Smith (1957) stated that they found no differences in AID between brain damaged and schizophrenic patients for two of the trials on which it was measured. In complete contradiction to these two studies, Spivack and Levine (1957) found that their brain damaged subjects reported *longer* AIDs than did a group of neurotic controls.

It is apparent that, while the theorists are in complete agreement, the small amount of evidence available is not. It was the purpose of this study to gather further data to help clarify the situation. Furthermore, if the theorists cited were correct, measuring the duration of the afterimage should increase the diagnostic efficiency of the test by identifying those patients whose brain injury does not reduce the afterimage to zero duration. These false negatives would be identifiable by their markedly reduced AID, even though they could report an afterimage.

Method

Subjects

An ongoing sample of all male veterans admitted to the Neurology and Psychiatry services of the Minneapolis Veterans Administration Hospital were tested within five days of their admission. This sampling procedure was adopted to control for chronicity of illness, eliminate patients undergoing shock therapy, and reduce the number of patients on a drug regimen. This method of selecting the sample also had the advantage of reducing errors which might have arisen had the experimenter had knowledge of the patient's diagnosis at the time of testing.

Of 129 cases admitted to the Neurology service 31 could not be tested: 3 expired, 17 were confined to their beds, 8 were too confused, 3 were discharged before testing. Seventeen additional cases were lost due to apparatus failure which occurred during the time the sample was being drawn. Of 50 admissions to the Psychiatry service, 5 could not be tested: 4 were confined to their beds, and 1 was discharged before testing.

Diagnosis

All diagnoses were final diagnoses arrived at by staff action. All cases having a diagnosis indicating central nervous system damage rostral to the foramen magnum were further classified by a boarded neurologist [2] who determined whether or not the injury included damage to the cerebral cortex; if it did not, case was classified under subcortical brain damage. The psychiatric patients were divided into those having a psychotic diagnosis and those having a nonpsychotic diagnosis. Three cases who were tested left the hospital before diagnostic work-up could be completed. They are listed as "equivocal" in the section to follow. The remaining patients, being neither brain damaged nor psychiatrically ill, constituted the normal hospitalized control sample.

Drugs

Each patient's medication at time of testing was noted from the order sheet in his hospital chart. Two physicians and a pharmacologist [3] then sorted the various drugs administered to patients in this study into four categories in accordance with their effect on the central nervous system within the range of therapeutic dosage. Drug Group 1 consisted of drugs having no effect on the central nervous system. Drug Group 2 contained all central nervous system depressants. Drug Group 3 included all central nervous system stimulants. Drug Group 4 consisted of any combination of drugs in Categories 2 or 3 which were being simultaneously administered to a patient.

Apparatus

An Archimedes spiral of 920° or 2½ turns about its center was constructed by painting a black spiral on a white disc 6 inches in diameter. The black and white lines were each ½ inch in width. The wheel was affixed to a reversible motor with a variable speed control which was set at 100 rpm.

Procedure

The *S* was seated 8 feet from the spiral. The illumination at the surface of the spiral was 16 foot candles, measured by a GE Exposuremeter, Model 8DW.

Two trials, one clockwise and one counterclockwise, were administered in accordance with Price and Deabler's (1955) instructions. Scoring was done in accordance with the rules established by Gilberstadt, Schein, and Rosen (1958). If the patient failed on one or both of these trials to report the negative afterimage, he was given two additional trials. At the end of stimulation on these trials he was told to look at an orange wooden block placed next to the apparatus and to describe what he saw. If he then reported a negative afterimage correctly on two successive trials (i.e., the block appeared to move, grow, or change), he was administered the

[2] Grateful acknowledgement is made to Zondal Miller, consulting neurologist, for sorting these patients.

[3] Grateful acknowledgement is made to Henry A. Johnson, Fred Miller, and A. Kleven for their assistance in classifying the drugs.

trials measuring duration. If he did not, he was given two further trials using a black circle drawn on white cardboard as the projection device. A patient who failed to report a negative afterimage on two successive trials of the six was considered to be diagnosed "brain damaged" by the test and no further testing was done.

As soon as two consecutive AIRs were given, the patient was instructed as follows:

> What you just saw is called an afterimage; i.e., the line (block, circle) seemed to be moving even though you knew it wasn't. People with normal vision see these afterimages, so in this sense your eyes are normal. Now I want to know how long the afterimage lasts. To find this out, I will run the wheel several times just as before. But this time when it stops I want you to look at this block (circle). If you notice anything change, tell me when the change stops. If it doesn't change in any way, tell me that, too. I have this very accurate stop watch here, so I can tell just how long your afterimage lasts. Any questions? All right now, keep your eye on the wheel until I turn it off. Then immediately look at the block (circle). As soon as the block (circle) comes to a *complete* stop, you say "Now."

Eight trials were administered in the following order:

Direction	C	CC	C	CC	C	CC
Stimulation Time	15 sec.	30 sec.	30 sec.	15 sec.	15 sec.	30 sec.
Direction	CC	C				
Stimulation Time	15 sec.	30 sec.				

(C = clockwise; CC = counterclockwise)

The variations in time and direction were introduced to discourage a fixed set to say "Now" at the same time following each stimulation, regardless of what may have been perceived.

At the conclusion of the fourth trial patients whose first afterimages had been passed off on the block were instructed to look at the circle following stimulation and vice versa. Whether a patient was exposed to the block or the circle on his first four trials was determined by a letter which had been written alternately on each scoring sheet prior to testing.

Two test objects were used (a) to provide variety in the situation to help keep the patient's attention and (b) to test the hypothesis that brain damaged individuals would have longer AIDs for the three-dimensional block than for the two-dimensional circle. The latter notion was suggested by the fact that brain damage is thought to reduce the individual's ability to deal with the world in an "abstract" fashion, thus the three-dimensional block might be more logically capable of motion than the two-dimensional circle.

Duration was measured by allowing a split-hundredth watch to run continuously from the start of the trial until the patient indicated the afterimage had stopped. The AID was determined by subtracting the stimulation time from the total time elapsed. This method reduced watch-reading and "snap-back" errors.

The interval between trials was maintained at 15 sec. from the time the patient indicated the afterimage had ceased. On the AIR trials, intertrial time was 30 sec.

Patients who failed entirely to report an afterimage or who reported less than 5 AIDs during the 8 trials measuring AID were not included in the calculations of the durations, on the grounds that their perception was so incomplete (or their ability to report so poor) that the time values they gave were of little value. The prior decision had been made to classify these patients as having been diagnosed brain-damaged by the test.

Results

The distribution of total AID scores (the sum of the durations reported for the eight timed trials) is shown in Fig. 1. Only those patients who reported five or more AIDs for the eight timed trials were included in the analysis. The number of cases who were excluded for this reason are listed in Table 1 under the heading "Less than Five Reports." Only 7 of the 123 cases failed to give an AIR in two consecutive pretrials, and only 3 of the remaining 116 cases did not report negative afterimages for at least five of the eight timed trials. The 3 equivocal (undiag-

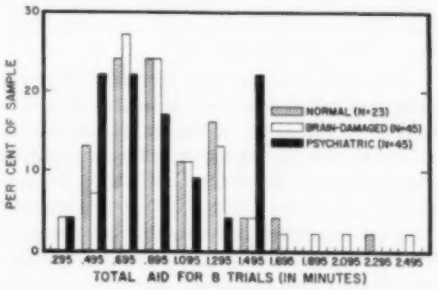

Fig. 1. Distribution of total AID (8 trials) by diagnostic groups.

TABLE 1

MEAN AGE, INTELLIGENCE, AND TOTAL AID (IN MINTUES) FOR ALL DIAGNOSED CASES

Diagnosis	*N*	Age	IQ	AID	Less than Five Reports
Brain damaged	53	42.78	106.36	.9911	8
Cortical	45	43.15	105.33	.9918	6
Subcortical	8	40.33	113.00	.9867	2
Psychiatric	46	37.85	110.64	.9758	1
Neurotic	34	37.06	110.76	.9309	1
Psychotic	12	40.00	110.33	1.0992	0
Hospitalized Normals	24	50.35	106.78	.9152	1
Total	123	42.35	108.15	.9696	10

nosed) cases are not included in the following analyses since they could not contribute any information to the diagnostic ability of the spiral test.

Table 1 summarizes the mean total AIDs, ages, and IQs for each diagnostic category. An analysis of variance of IQs indicated that the groups did not differ significantly on this variable. The product-moment correlation between total AID and IQ was .034, which was not significantly different from zero. The analysis of variance of age, however, did show that the diagnostic groups differed significantly ($p < .01$) on this variable. The product-moment correlation between total AID and age was found to be $-.092$ ($p > .05$), thereby making a covariance adjustment for age unnecessary.

The analysis of variance of differences in total AID score by diagnostic groups confirmed what was apparent from inspection of Fig. 1. The three diagnostic categories do not differ significantly in the reported duration of the afterimage.

An analysis of variance in which the groups were further divided by the drugs they were taking at the time of testing was performed to see if this added factor might account for the lack of a significant difference between diagnostic groups. Since only Drug Groups 1 and 2 were represented in each of the diagnostic categories, only these two groups were included in the analysis. None of the sources of variation contributed a mean square greater than that for individuals (within variance) which served as the error term (McNemar, 1955, pp. 304–306).

The lack of differences between diagnostic groups on total AID as a function of drugs might have resulted from too gross a grouping of the various depressant medications. Drug Group 2 was, therefore, further subdivided into the more potent drugs (sparine, thorazine, trilafon, vesprin, and compazine) and the less potent drug, meprobamate. Only the psychiatric group was included in this analysis, because there was no certain way to assess the effects of the anticonvulsant medication which provided the main group of drugs prescribed for the brain damaged patients. The latter patients' AID scores were separately analyzed.

The analysis of variance of AID scores for the psychiatric patients, comparing those who received central nervous system depressants with those receiving meprobamate and those receiving no centrally acting drug, yielded a nonsignificant *F*. The AID scores for the 28 brain damaged patients diagnosed as having seizures were compared on the basis of whether or not they were on anticonvulsant medication at the time of testing. Again, the analysis of variance failed to indicate a significant difference attributable to this factor.

The diagnostic groups had differed on yet another variable, order of presentation of the test object; i.e., whether the block or circle was used as a test object for the first four duration trials. Table 2 presents the results of the analysis of variance of order of presen-

tation by diagnostic groups. Again, diagnosis was not significant, while order of presentation of the test object did significantly affect AID. The circle tended to be associated with longer duration times than the block for all diagnostic groups. For those *Ss* to whom the circle was presented first the mean ratio of their first four to their last four duration times was .76; for those receiving the opposite presentation the mean ratio was .96. Therefore, when the circle was used as a test object on the last four trials, there was a mean decrement of only 4% in AID reported; while, when the block was used on the last four trials, the mean decrement was 24%.

The possibility remained that order of presentation of the test objects might have affected the total AID; and, since the number in each diagnostic category receiving the two orders of presentation were not equally distributed, this might have contributed to the failure of the AID to distinguish between the groups. However, an analysis of variance showed that the total AID did not differ significantly between those presented with the block for the first four trials as against those presented with the circle for the first four trials.

Another factor which might have been of value in distinguishing between the diagnostic groups was stimulation time. A briefer stimulation time might reduce AID more for brain damaged than for other patients. The analysis of variance for each of the stimulation times showed that neither the 15-sec. nor 30-sec. period was successful in significantly differentiating the diagnostic groups.

Stimulation time itself, however, was significantly related to the AID. The mean total AID following 15-sec. stimulation was .4306 min. and following 30-sec. stimulation was .5388 min. The critical ratio of this difference was significant at less than the .001 level.

TABLE 2

ANALYSIS OF VARIANCE OF EFFECT OF ORDER OF TEST PRESENTATION BY DIAGNOSTIC GROUPS ON TOTAL AID

Source	*df*	*F*
Order	1	9.9396*
Diagnosis	4	.5202
Interaction	4	.0896
Individuals	101	(1,096.77)[a]

* Significant at less than .01 level.
[a] Error variance.

TABLE 3

DISTRIBUTION OF NUMBER OF FAILURES TO REPORT AFTERIMAGES

Frequency of Failure	Brain Damaged ($N=53$)	Psychiatric ($N=46$)	Normal ($N=24$)
6+	5	1	1
5	0	0	0
4	3	0	0
3	1	1	0
2	1	1	0
1	10	1	2
0	33	42	21

Turning from the analysis of AID, the number of zero responses was examined for possible diagnostic significance. The distribution of number of times that patients reported no afterimage is given in Table 3. All cases except the three having no diagnosis are listed. The category "6+" contains those patients who either reported no afterimage for six or more trials determining the duration of the afterimage or who did not report two consecutive afterimages in the pretrials. When the number of zero reports was dichotomized between none and one or more, a chi square of 13.69 was obtained ($p < .01$, $df = 2$). In percentage terms, 37.7% of the brain damaged cases reported one or more failures to perceive the afterimage on either the initial or duration trials. This figure compares with 8.7% of the psychiatric group and 12.5% of the hospitalized normals who failed to report an afterimage one or more times.

DISCUSSION

The results of this investigation fail to support any of the theories which reason that brain damage causes a decrease in cortical efficiency or an increase in cortical inhibition —either of which would result in the reduced duration of the afterimage. The failure of the results to satisfy the predictions of such theories cannot be laid to differences between the criterion groups in age, intelligence, drug status, or gross differences in diagnosis (i.e., psychosis vs. neurosis, cortical vs. subcortical

brain damage). Neither age nor IQ correlated significantly with AID, nor did depressant drugs have a different effect on AID than no drug when patients with similar diagnoses were compared.

Eysenck's et al. (1957) finding that depressant drugs would foreshorten the AID was not confirmed by this study. However, their experimental *Ss* were without any known pathology, and the drug's effect on the *S*'s behavior was measured shortly after a single large dose had been administered. The patients in this investigation were selected because they were ill in one way or another, and those who received depressant drugs were given them in doses spaced over a period of time, with the elapsed time between drug ingestion and test administration being uncontrolled. In addition, the depressant drug used by Eysenck et al. (1957) was a barbiturate, whereas in this investigation almost all of the patients in the depressant category were given ataractic or anticonvulsant medication. That these factors are important in assessing the effects of drugs is attested to by Lehmann and Csank (1957) who write:

> It is probable that the same drugs would produce differing results after the subjects had taken them regularly for a time because of tolerance or sensitization . . . . It is also possible and even probable that severely disturbed patients will show reactions that differ considerably from the pattern observed in healthy control subjects (p. 231).

According to Shapiro's (1954) theory of cortical inhibition, the differences between the diagnostic groups should have been most marked following the longer stimulation period, since cortical inhibition would be aroused more quickly in the brain injured, depriving them of the increase in AID which would follow longer stimulation. Instead, it was found that differences in diagnostic status did not contribute to the significant difference in AID found between the 15-sec. and 30-sec. stimulation periods. The fact that amount of stimulation did affect AID, but not differently for the diagnostic groups, leads to a further question as to the adequacy of cortical inhibition and efficiency theories as now proposed.

The average duration following 15-sec. stimulation was 6.5 sec. and following the 30-sec. stimulation was 8.0 sec. Considering this brief period of time, the possibility of a sizeable error of measurement caused by individual differences in interpretation and execution of the task is manifest. Some *Ss* might be expected to delay slightly before reporting cessation of the afterimage, in order to be certain that they were reporting accurately. Others might report cessation even before it had occurred in hopes of bringing the testing to a close more quickly. With the duration period per trial being so small, idiosyncrasies in reporting would have a substantial effect on the results. With this thought in mind, it is interesting to note that the correlation between the AID following the four 15-sec. trials with those following the four 30-sec. trials was .838, indicating a fairly high degree of consistency in reporting the afterimage. It would seem that whatever factors underlie the reporting of the AID they are relatively constant for each individual.

That there was difficulty in interpreting the task is attested to by the fact that the AID associated with the circle as test object was significantly longer than that associated with the block. This finding suggests that two afterimages were involved, one resulting from stimulation by the rotating spiral and one due to the contrast of the black circle on the white ground. It is well established that prolonged exposure to a highly contrasted, stationary stimulus will result in a negative afterimage. It is likely that most of the patients interpreted this afterimage as also being a part of the same phenomenon they were requested to report. The orange block against the neutral ground either did not produce this type of afterimage or did so to a lesser extent.

The conclusion that duration of the afterimage does not differ between brain damaged and other diagnostic groups is consistent with the previously reported study by Page et al. (1957) and inconsistent with that of Spivack and Levine (1957). The former study involved adults, while the latter sampled adolescents. Page et al. unfortunately did not report their data on AID, other than to note that it was not significant. Spivack and Levine, however, report the mean durations for

their brain damaged and control groups, which were 20.2 and 14.8 secs., respectively. Either group has substantially longer AIDs than the 8.0-sec. average for the four 30-sec. stimulation periods of this study. One explanation for this difference can be found in the fact that they seated their *Ss* five feet from the spiral. At this distance, more rod vision is involved which should *decrease* the AID (Granit, 1928; Wohlgemuth, 1911). At the same time, however, it has been noted (Granit, 1927, 1928) that the duration of an afterimage produced in the periphery of the retina is less distinct and more difficult to measure even by highly trained *Ss*. It is possible, then, that differences in sampling and in procedure may account for the inconsistency of the results found by Spivack and Levine and those obtained in this study.

One further limitation of this study as well as the two mentioned above is that differences, or lack of differences, in AID found between brain damaged and other groups may be the result of slowed reaction time in the brain damaged. The brain damaged patient may perceive the afterimage as cessating sooner but be slower to report this fact. A method which circumvents verbal report would be necessary to resolve this question. One suggestion, currently being investigated, would be to use electroencephalographic recordings of occipital potentials as the measure of afterimage duration. If the stimulation by a rotating spiral results in changes in the occipital cortex, then it is likely that this change will result in a suppression of occipital alpha for a period following stimulation which is closely related to the perceived duration of the afterimage. The use of such a measure would circumvent verbal report and permit a more direct study of this phenomenon.

Turning to another explanation of these findings, it must be noted that, while AID did not differentiate the diagnostic groups, AIR did significantly separate them. Significantly more brain damaged individuals failed to report a negative afterimage one or more times than did the psychiatric or normal groups. Spitz (1958) has argued that failure to report the illusion is evidence for lack of an "abstract attitude." Certainly there is considerable ambiguity in the question: What does the line appear to be doing now? It is conceivable that some patients structure this query as referring to the objective condition of the spiral when the experimenter has turned it off. In other cases, failure to report the afterimage may be due to a suspicion on the part of the patient that the perception of the illusion is indicative of some kind of pathology. Thus, denial of the afterimage may be a symptom related to anosognosia, another symptom fairly common in brain damaged individuals (Tobis, Lowenthal, & Maringer, 1957).

Since only 10 of the 123 patients tested did not report a negative afterimage on half or more of the trials, there is no question that the use of a test object increases the likelihood of AIR. That this is true suggests that some of the failures to report the afterimage may be due to confusion about the task. We may look upon this finding as lending further support to the alternative hypothesis that brain damage does not effect the perception of the afterimage, though it may differentially effect the frequency with which this phenomenon is reported.

Summary

An ongoing sample of male veterans admitted to the neurology and psychiatry services of a general hospital were tested with the Archimedes spiral. In addition to determining whether or not the patient could report the negative afterimage, the duration of the negative afterimage was also measured. To decrease the ambiguity of the task, two test objects were used on which the patient was instructed to pass off his afterimage. This technique greatly increased the number of *Ss* who reported the afterimage. Neither age, intelligence, drugs, nor diagnosis differentiated the patients with regard to the duration of the afterimage. However, when the number of failures to report the afterimage on any of the 10 trials was used, it was found that the brain damaged patients could be separated from the psychiatric and hospitalized normal patients at a highly significant level. It was suggested that these findings call into serious question theories of brain functioning which depend upon a cortical effi-

ciency or cortical inhibition hypothesis. The alternative explanation of studies which indicate that brain damaged patients have greater difficulty in reporting the spiral afterimage seems to be that the brain damaged person is more readily confused by the task. The data from this investigation suggest that brain damaged patients perceive the afterimage as frequently as other patients, but they fail to report their perception more frequently.

REFERENCES

BERGER, D., EVERSON, R., RUTLEDGE, L., & KOSKOFF, Y. D. The spiral aftereffect in a neurological setting. *J. consult. Psychol.*, 1958, **22**, 249–258.

EYSENCK, H. J., HOLLAND, H., & TROUTON, D. S. Drugs and personality: III. The effects of stimulant and depressant drugs on visual aftereffects. *J. ment. Sci.*, 1957, **103**, 650–655.

GALLESE, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.

GILBERSTADT, H., SCHEIN, J. D., & ROSEN, A. Further evaluation of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, **22**, 243–248.

GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, **22**, 153–157.

GRANIT, R. Über eine Hemmung der Zapfenfunktion durch Stabchenerregung beim Bewagungsnachbild. *Z. Sinnesphysiol.*, 1927, **58**, 95–99.

GRANIT, R. On inhibition in the aftereffect of seen movement. *Brit. J. Psychol.*, 1928, **19**, 147–157.

KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.

LEHMAN, H. E., & CSANK, J. Differential screening of phrenotropic agents in man: Psychophysiologic test data. *J. clin. exp. Psychopath.*, 1957, **18**, 222–235.

MCNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.

PHILBRICK, EMILY B. The validity of the spiral aftereffect as a clinical tool for diagnosis of organic brain pathology. *J. consult. Psychol.*, 1959, **23**, 39–43.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

SAUCER, R. T. The nature of perceptual processes. *Science*, 1953, **117**, 556–558.

SAUCER, R. T., & DEABLER, H. L. Perception of apparent motion in organics and schizophrenics. *J. consult. Psychol.*, 1956, **20**, 385–389.

SHAPIRO, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *Brit. J. Psychol.*, 1954, **45**, 58–67.

SPITZ, H. H. The present status of the Kohler-Wallach theory of satiation. *Psychol. Bull.*, 1958, **55**, 1–28.

SPIVACK, G., & LEVINE, M. The spiral aftereffect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, **25**, 1–11.

TOBIS, J. S., LOWENTHAL, M., & MARINGER, S. Evaluation and management of the brain-damaged patient. *J. Amer. Med. Ass.*, 1955, **65**, 2035–2041.

WOHLGEMUTH, A. On the after-effect of seen movement. *Brit. J. Psychol., Monogr. Suppl.*, 1911, **1**, 1–117.

(Received July 29, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 307-310

A STUDY OF IDENTIFICATION IN MALE HOMOSEXUALS[1]

JUDY CHANG AND JACK BLOCK
University of California, Berkeley

In discussing the etiology of male homosexuality, Freud (1952) has written:

> The typical process . . . [in homosexuality] is that a few years after the termination of puberty, the young man, who until this time has been strongly fixated to his mother, turns in his course, identifies himself with his mother, and looks about for love-objects in whom he can re-discover himself, and whom he wishes to love as his mother loved him (p. 240).

According to Freudian theory, therefore, homosexuality in males is a manifestation of an identification with the mother figure rather than the usual identification with a father figure. As a consequence of his twisted identification, the homosexual seeks narcissistic rather than anaclitic love-objects.

The purpose of the present study is to test a consequence of this notion, namely, that male homosexuality involves both a turning *toward* the mother as a figure with which to identify *and* the turning *away* from the father as an identification figure. Specifically, the major hypothesis of this investigation was that frankly homosexual males have a pattern of parental identification such that they are more strongly identified with their mothers and, conjointly, more strongly disidentified with their fathers than are males in a relevant contrast group. This conjoint hypothesis of course can be further partitioned into two separate hypotheses of interest, permitting comparison of the homosexual and nonhomosexual subjects (*Ss*) with regard to their *absolute* degree of identification with mother and with father figures. The two subhypotheses that follow from the major hypothesis are: (*a*) overt homosexual males, when compared to nonhomosexuals, will show a stronger identification with their mothers, and (*b*) overt homosexual males will show less identification with their fathers than will members of the control group.

The meaning of the Freudian term identification has been the subject of some controversy (Sanford, 1955; Stoke, 1950). For research purposes, various operational translations of the concept have been employed. The operational meaning of identification adopted for the purposes of the present study is based upon the following statement by Freud (1933): ". . . when a boy identifies with his father he wants to *be like* his father" (p. 90). According to Freud, at least in this late statement by him on the matter, identification with an object is said to have occurred when an individual strives to take on attributes which are those of that object as the individual perceives that object; this object is no more and no less than one's ego ideal—the object ". . . towards which it . . . [the ego] strives" (Freud, 1933, p. 92). In this study, then, a measure of the extent of identification with one's mother or father following directly from this conceptual definition was derived by finding the degree of correspondence between the attributes of one's ideal self, i.e., the person one would like to be, and the attributes of one's mother or father, as described by the identifier. The higher the degree of correspondence between one's ego ideal and one's parent, the greater the identification with that parent.

METHOD

Two groups of *Ss* were used. One group, the experimental group, consisted of 20 overt homosexual males. These *Ss* were engaging in homosexual practices, i.e., most of them were involved in homosexual marriages. None of these *Ss* were receiving or, with the exception of one, had ever received psychotherapy. Thus, all of these *Ss* appeared to

[1] This investigation, part of a larger research project, was supported by research grant M-1078 from the National Institute of Mental Health of the United States Public Health Service.

have accepted their homosexuality, a most important and perhaps unusual property of this sample. Their cooperation to serve as *Ss* was obtained through personal and necessarily confidential channels. These *Ss* were aware that they had been asked to participate in the experiment because of their homosexual practices but had not been informed of the purpose of the experiment. A second group of 20 males from the same urban community and equivalent with respect to age and education was used as a contrast group. Eight of these *Ss* were married, and twelve were still attending college. The mean age of the homosexual males was 31.9 years with a range from 22 to 39, that of the contrast group, 29.7 years with a range from 22 to 39. The mean educational level was 14.5 years for the homosexual group and 14.9 years for the contrast group.

The measure of identification with parent was obtained through the use of a list of 79 adjectives selected to permit a comprehensive personality description.[2] Each *S* was instructed to describe his "ideal self" by the use of the adjectives, marking an X if he felt that the adjective was particularly characteristic of his ideal self or an O if it was particularly uncharacteristic. The *S* was allowed to make only and exactly 30 X's and 30 O's in order to avoid the confusions engendered by response sets (Cronbach, 1950). Following this description, *S* was required to describe, in the same manner, "your mother," "your father," and "yourself (as you see yourself)." In order to facilitate independence of the several descriptions, each object description was made on a separate page of the experimental booklet.

Of 35 booklets circulated into the available homosexual population, only 22 were returned. Of these, only 20 were found to be usable in the sense of being complete or correctly filled out. Data for the contrast group then was collected by distributing booklets among individuals who were similar to the homosexual group with respect to age and educational level. Of 31 individuals requested to participate in the experiment, only 21 were willing to do so. One of these 21 control *Ss* was arbitrarily eliminated from the subsequent data analysis in order to equalize the sample size. In view of the fractional returns from the populations sampled, the possibility of some kind of selective bias in the results should, of course, be recognized.

For each *S*, a mother-identification score and a father-identification score was obtained by comparing his description of his ideal self with that of his mother and with that of his father, respectively. For a given adjective, a score of one was given if it was marked X, O, or left unmarked on both the ideal-self- and parent descriptions. An identification score was obtained by summing the number of these correspondences between the ideal-self-description and the parent description. The separate mother- and father-identification scores thus derived were used to test the two subhypotheses. For the major hypothesis, that male homosexuality represents simultaneously an idealization of mother together with a disparagement of father, a derived score was obtained by subtracting for each *S* his father-identification score from his mother-identification score. The score distributions for the mother-identification and father-identification variables had quite equivalent means and standard deviations (40.7 and 11.3; 41.0 and 11.5, respectively). Accordingly, it was legitimate to calculate the derived score from the original data distributions without going to the trouble of standardizing the separate score distributions. Thus, this derived score would be high if an *S*'s mother-identification score was high and/or his father-identification score was low. In addition, a self-acceptance score was obtained by comparing the ideal-self-description with the self-description, scoring for correspondence in the manner described above. These four scores provided the basic data for analysis.

For each comparison of interest, the Mann-Whitney Test (1947) was used to calculate the significance of the difference between the scores obtained by the experimental and control groups. With the exception of the analysis of self-acceptance scores, all the statistical results reported in this study are based on one-tailed tests of significance because the direction of difference was predicted in advance.

Results

For the homosexual group, the derived scores expressing the primary variable, i.e., distance between mother and father on *S*'s identification continuum, are significantly greater (.001) than the derived scores of the contrast group. Some indication of the separation between the two groups on this variable is provided by the finding that 14 of the 20 homosexuals have positive scores where only 4 of the 20 nonhomosexuals are on the positive side of zero. The primary hypothesis, that overtly homosexual males are relatively more strongly identified with their mothers than with their fathers, is therefore supported by this finding.

When the primary variable is partitioned into its components (the mother-identification score and the father-identification score) and separately analyzed, a better perspective is gained on the nature of the preceding finding. When a comparison of the mother-identification scores of the two groups is made, the homosexual *Ss* show a significantly greater degree of identification with their mothers. The difference between the sums of ranks of mother-identification scores of the two groups of *Ss* is significant at the .03 level. An analysis of the father-identification scores of the two groups of *Ss* shows a not quite

[2] A copy of this list of adjectives is available by writing to the authors.

significant but still appreciable trend in the direction of less identification with father by the homosexual group. The one-tailed test results in a p value of .07, almost reaching the conventional level of significance. From these separate and rather equivalent findings, it is clear that the striking separation between the two groups in the test of the first hypothesis is a *joint* reflection of the two constituent scores rather than being dominated by but one of these. This kind of result is both a subtle and a more interesting one.

Self-Acceptance

The homosexual and control groups did not differ significantly in their extent of self-acceptance, where self-acceptance is defined as the degree of correspondence between the perceived self and ideal self. It should be especially noted that there is no perceptible trend for the contrast group to be relatively more self-accepting, as indicated by the two-tailed p value of .64 for the obtained difference between the sums of ranks. This finding is a rather interesting one for it calls attention to the special nature of this homosexual sample as one *not* psychiatrically disturbed and apparently well-accepting of their homosexual activities.

Analysis of Ideal-Self-Descriptions

The results in the preceding analyses might have been due to differences between the kinds of ego ideals of the homosexual and control groups. For this reason, an item analysis of the adjective descriptions of the ideal selves of the two groups of *Ss* was done. Only 2 of the 79 adjectives discriminated (at the .05 level) between the ideal-self-descriptions of the two groups. The conclusion is made, therefore, that the differential valuations by the homosexual *Ss* and control *Ss* of their parents are *not* due to different ego ideals. For the sake of completeness, it may be recorded that nonhomosexual *Ss* considered "sympathetic" as an attribute of their ideal self more frequently than did homosexual *Ss*. Homosexual *Ss*, more than control *Ss*, tended to deny the characteristic "dependent" as an attribute of the ideal self. No additional adjectives were found to discriminate when the threshold of significance was lowered to the .10 level.

Analysis of Self-Descriptions

Since the homosexual and control *Ss* did not differ in their degree of self-acceptance and in their ideal-self-descriptions, one might expect that they would also not differ in their self-perceptions. But if this expectation were to be realized, then some of our preceding findings become questionable. The reasoning behind this remark is that homosexuals and normals are quite obviously rather different sorts of people and, consequently, these differences should, if only on a priori grounds, be manifest if a reasonably sensitive instrument is employed. Failure to find these differences would be presumptive evidence of the technique's inadequacy, a conclusion that might also explain away the failure to find differences with respect to self-acceptance and ego ideals. On the other hand, it is possible, despite the failure of the self-acceptance and ideal-self-analyses to show differences, for the self-description analyses to separate the groups. This possibility arises because the self-acceptance for neither group is so high as to require (because of equivalent ideal selves) that the content of the self-descriptions be equivalent. With moderate correlations of self- with ideal-self-descriptions, the different dimensions of self-description are free to operate and, hence, be seen.

Item analysis of the self-descriptions of homosexual and control *Ss* finds six adjectives discriminating at the .05 level, with an additional six adjectives meeting the .10 criterion —results not accountable by a proper statement of chance significance.[3] Moreover, the discriminating adjectives are consonant with the concept of the homosexual male. Compared to control *Ss*, the homosexual male sees himself as "affected, dependent, determined, personally charming, restless, and *not* tact-

[3] By reference to "a proper statement of chance significance," we are alluding to a recently developed recognition that conventional expectations of chance significance have been appreciably wrong rather often. In the situation where the response distributions of both groups being compared are highly skewed (or in the dichotomous case, are extremely unbalanced as in the present study), at the, for example, .05 level, many less than 5% of the findings will be significant when chance alone is operating. The logic underlying this assertion and some empirical findings are detailed in a manuscript available from the second author.

less" (all of these significant at the .05 level). The picture can be extended by noting that the homosexual tends to see himself as "impulsive" but not "frank" or "easily embarrassed" or "ambitious." Our homosexual males are less often "friendless or introspective or self-aware" than the control *Ss*, all of these last differences reaching the .10 level.

Although the discriminating characteristics cluster and portray someone with the personality traits usually associated with homosexuality in this culture, it is especially noteworthy that such adjectives as "confident, psychologically secure, relaxed, dissatisfied, helpless, worried" and the like failed to discriminate between the self-perceptions of the two groups.

Discussion

The major hypothesis of this study, that frankly homosexual males more than control males show a positive identification with their mothers together with a disidentification with their fathers, has been supported by the results of this study. Subsidiary hypotheses have also been supported. What now shall we make of this finding?

Clearly, no one study is able to test the Freudian hypothesis, in all its complexities, about homosexuality. The present research does, however, in its limited way, support the psychoanalytic position. Additional studies, by the present and alternative methods, are required to investigate the multifaceted Freudian theory that the relatively strong identification with mother found in homosexual males has been preceded by and is the result of a strong fixation upon the mother, that the starting point of homosexuality comes quite discretely in adolescence, and that homosexuality represents a narcissistic object-choice. Such a research represents an undertaking of a much larger dimension than the present one.

With respect to the present study, the unusual nature of the experimental sample used —self-accepting homosexuals—should especially be noted. Our results may well be an expression of the special character of this group. In more conflicted homosexuals, where identifications have been less decisively made, the presently obtained relationships can be expected to exist less strongly. One should find, among individuals with more ambiguity of identification, that there is less self-acceptance and more personal despair than was observed among the effectively functioning and well-received homosexuals studied in the present research.

Summary

The present study was concerned with a test of the Freudian notion that homosexuality in the male is based upon an over-identification with the mother figure together with an under-identification with the father figure. Two groups of *Ss*, equivalent with respect to age, education, and socioeconomic level, were used: a group of 20 males engaging, apparently successfully, in homosexual practices, and a group of nonhomosexual males. Each *S* described his ideal self, his mother, his father, and himself, by the use of a list of adjectives. On the basis of an "ego ideal" conception of identification, various identification scores and a self-acceptance score were obtained for each *S*, by comparing his description of his ideal self with that of the other individuals he described.

As tested, the hypothesis was supported for it was found that overt homosexual males were more identified with their mothers and more disidentified with their fathers than were the nonhomosexual males with whom they were compared. In addition, it was found that the two groups of *Ss* did not differ significantly in their degree of self-acceptance nor in regard to the kind of ego ideal toward which they aspired. The findings were briefly discussed.

REFERENCES

Cronbach, L. J. Further evidence on response-sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

Freud, S. *New introductory lectures on psychoanalysis.* New York: Norton, 1933.

Freud, S. *Collected papers.* Vol. 2. London: Hogarth, 1952.

Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

Sanford, N. The dynamics of identification. *Psychol. Rev.*, 1955, **62**, 106–118.

Stoke, S. M. An inquiry into the concept of identification. *J. genet. Psychol.*, 1950, **76**, 163–189.

(Received July 29, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 311-315

THE AFTEREFFECT OF SEEN MOVEMENT AND BRAIN DAMAGE

R. H. DAY
University of Sydney

If a moving stimulus object which has been visually fixated for about one minute is stopped suddenly, there occurs a characteristic illusory movement in a direction opposite to that of the real movement. This aftereffect of seen movement is experienced following fixation of various types of moving stimulus objects including a rotating spiral pattern, a sectored disc, or a pattern of black and white bands (waterfall illusion). The phenomenon is not restricted to the stimulus object but can be observed in the field of the closed eye and on other textured surfaces. If one eye is used during fixation and that eye closed on the cessation of real movement, the aftereffect can be observed using the nonstimulated eye.[1]

Recent studies have drawn attention to a relationship between this aftereffect and brain damage. So far only the rotating spiral pattern has been used as a stimulus object (Aaronson, 1958; Davids, Goldberg, & Laufer, 1957; Freeman & Josey, 1949; Gallese, 1956; Garrett, Price, & Deabler, 1957; Holland & Beech, 1958; Price & Deabler, 1955; Spivack & Levine, 1957). Results from these studies have indicated that organically impaired subjects (*Ss*) can be discriminated from normal or functionally disordered controls with a high degree of certainty. Generally it has been found that, whereas normal individuals experience a vivid and unmistakeable aftereffect, cases of brain damage either report no aftereffect or one which is reduced in its duration or vividness. Some studies, however, have not confirmed this and have reported that little or no differentiation can be made between brain damage and normal controls (Goldberg & Smith, 1958; Johnson, Bauer, & Brown, 1959; Page, Rakita, Kaplan, & Smith, 1957; Philbrick, 1959; Pickersgill & Jeeves, 1958).

It is intended here to direct attention to issues which appear to have been overlooked and which have an important bearing on research in this area. It is believed that failure to consider certain behavioral outcomes of brain damage as well as insufficient attention to methodology and sampling have resulted in the lack of agreement between the results from independent studies.

Visual Fixation and Movement Aftereffects

The aftereffect of seen movement occurs when the eye is fixated on a stationary point during the period of stimulation by the moving pattern (Holland & Beech, 1958; Wohlgemuth, 1911). If the eye follows the moving pattern, little or no aftereffect occurs. In order to confirm this 12 normal *Ss* were required to follow one of the sectors on a black and white sectored disc on one trial and to fixate the center of the disc on another. During the following trial eight *Ss* failed to report an aftereffect and four reported a slight but brief aftereffect. All *Ss* experienced a strong aftereffect (apparent movement in the reverse direction) following fixation of the center of the disc.

When the eye follows a point on a moving stimulus pattern, the retinal image of the pattern either remains stationary or moves less relative to the retina. When a fixed point such as the center of the disc is fixated, in retinal image moves with respect to the retina. An essential condition, therefore, for the occurrence of a clear aftereffect of seen movement

[1] The apparent interocular transfer of an aftereffect from the stimulated to the nonstimulated eye cannot be interpreted as evidence for its central (cortical) locus (Day, 1958).

is motion of the retinal image of the stimulus pattern. This condition is achieved by steady fixation of a stationary point in or near the stimulus object.

Fixation and Brain Damage

In an investigation of the cortical integration of ocular movements it has been shown that, whereas some organics experienced difficulty in voluntarily changing the point of fixation, their eyes followed a moving object involuntarily (Holmes, 1938). In an account of the effects of frontal brain damage on eye movement and fixation Holmes has stated:

> This condition at first shows all the appearance of loss of paralysis of movement, but that this is not present becomes obvious when the subject is asked to follow with his eyes a slowly moving object; then, providing the rate of movement is uniform and not too fast a full range of deviation may be obtained . . . when the patient looks at a rotating drum on which lines or figures are displayed the eyes also deviate in the direction in which the figures move . . . We therefore find that though voluntary effort fails to produce movement the eyes follow a moving point the images of which fall on the retina (p. 106).

In discussing the work of Holmes (1938) Sherrington (1940) has presented a similar view:

> But if asked to follow with his eyes a slowly moving object, his eyes do so. That is, though willed effort fails to produce the act, the eyes follow 'of themselves' a moving point whose image is fixated by them. The eyes remain anchored to that object by the 'fixation reflex.' Those afflicted with this condition when they want to look at something fresh, close the eyes for a moment, or toss the head, or place a hand between the eyes and the object fixing their gaze. By breaking the gaze a moment they free themselves from the fixation reflex; though by willed effort they cannot break it (p. 226).

Thus in those cases in which voluntary eye movements and visual fixation have become difficult or unattainable, but in which optical fixation reflexes are preserved, the eyes remain fixated on the moving object. This symptom of frontal brain damage has been called "following movement" by Walsh (1957). It would be expected then that the moving elements of a stimulus pattern would tend to be followed, rather than a stationary point fixated. It has been pointed out that this method of observation does not result in a persistent or vivid aftereffect. It seems probable that the absence, weakness, or short duration of the aftereffect of seen movement observed in certain cases of brain damage derives in some measure from an inability to meet the conditions of visual fixation necessary for its occurrence. Even if fixation of a stationary point is achieved during part of the period of stimulation, the aftereffect would be reduced as compared with that resulting from steady fixation during the full period of stimulation. The length of the period of stimulation is an important determinant of the aftereffect (Wohlgemuth, 1911).

A possible relationship between the well known distractability of organically impaired *Ss* and reduced aftereffects of seen movement has already been pointed out (Holland & Beech, 1958). It has been suggested that such distractability could result in inadequate fixation and thus contribute to the weakness of the aftereffect. The view proposed here, however, is that the weakness of the aftereffect is based on an inability on the part of the organic *S* to fixate rather than on a tendency to inattention.

Movement Aftereffect and the Locus of Brain Damage

The control of eye movements is not confined to a single cortical center. Stimulation of the frontal region (Brodman's Area 8), the inner aspects of the occipital lobes and temporoparietal centers produces eye movements. Penfield and Rasmussen (1950) located an eye movement (turning) center at the junction of the "face" and "hand" areas in the sensorimotor sequence but extended anteriorly with respect to the rest of the sensorimotor strip. Eye turning has also been elicited in some cases by stimulation of the temporal and parietal lobes (Penfield & Boldrey, 1937).

The most marked disturbances of voluntary fixation occur when the frontal region is affected, especially when there is involvement of the frontal projection fibers in the internal capsule. Disturbances of fixation also occur when the fibers in the anterior region of the midbrain are affected (Holmes, 1938).

It would be expected, therefore, that the degree to which the aftereffect of seen movement is affected would depend upon the nature, locus, and extent of brain damage, the most marked deterioration occurring with damage to those centers controlling voluntary fixation. There have been no studies so far in which the locus and magnitude of damage have been considered in relation to the aftereffect. The most common practice has been to compare the aftereffect reported by a heterogeneous organic group with that reported by normal and functionally disordered controls. Since the constitution of the brain damaged groups has varied widely between the numerous investigations, it is not surprising that inconsistencies between results have occurred.

It seems probable that the discrepancies between independently collected data is due in part to a failure to consider the locus, chronicity, and extent of brain damage. Considering the wide variety of behavioral symptoms manifested by such diverse conditions as epilepsy, traumatic brain injury, Parkinson's disease, and postsurgical conditions there is little justification for regarding brain damage as a single category characterized by a common experience (or lack of it) following visual stimulation. It is suggested here that the widely differing results derive from the varying effects of different types of brain damage on ability to fixate under the conditions employed.

Scoring Methods

Two methods for scoring the aftereffect of seen movement have been used. The simplest method is that of scoring the reported occurrence and nonoccurrence of the aftereffect 0 and 1, respectively, and an equivocal report 0.5 (Freeman & Josey, 1949).[2] A variant of this method is to score nonoccurrence 0, occurrence 1, and to score 0.5 when a single phenomenal feature only (contraction-expansion, change in size or approach–recession) is reported (Garrett et al., 1957; Philbrick, 1959; Price & Deabler, 1955). A second method requires the *S* to report the duration of the aftereffect (Holland & Beech, 1958; Page et al., 1957; Spivack & Levine, 1957).

It has been found that, whereas some brain damaged *S*s fail to report the aftereffect spontaneously, they have reported it when the nature of the phenomenon has been described to them. This has been most noticeable when a rotating black and white sectored disc has been used as the stimulus object. The aftereffect from this pattern (apparent reversal of movement) is much less vivid than from a spiral pattern. It has been observed also that organically impaired *S*s often report a single feature spontaneously but will describe another feature following a description of it.

It is conceivable that with respect to both the presence or absence of the aftereffect and the number of reported features the differences between organic and normal *S*s derive from the adequacy with which they report. If the duration or persistence of the aftereffect were used as the score after it has been established that the *S* experiences the phenomenon, the reliability of the scores could be readily established. It has been found that duration can be reported with confidence by most organic *S*s with the exception of those who exhibit marked deterioration and confusion.

Discussion

Two explanations of the absence or weakness of the aftereffect of seen movement reported by organic *S*s have been suggested. The first postulates a synthesizing or integrating process in the CNS responsible for movement perception. It is argued that this process is either absent or adversely affected by brain damage (Price & Deabler, 1955). This view is based upon a recent theory of motion perception (Saucer, 1953, 1954). A second explanation derives from a concept of central nervous inhibition. The effect of brain damage is to intensify cortical inhibitory processes and thus to weaken the aftereffect (Holland & Beech, 1958; Price & Deabler, 1955). Each of these explanations emphasizes the role of central nervous mechanisms as determinants of the aftereffect of seen movement.

[2] Whereas Freeman and Josey (1949) score nonoccurrence of the aftereffect as 1 and occurrence 0, other investigators have scored 1 for occurrence and 0 for nonoccurrence. The latter would appear to be a more meaningful method.

The view expressed here is that the failure of some organic *Ss* to experience the aftereffect from a moving pattern, or to experience it with less vividness than normal *Ss*, is due to a tendency to visually follow the elements of a moving pattern. This tendency is consequent upon damage to those parts of the brain initiating and controlling voluntary fixation and a replacement of these voluntary activities by reflex movements controlled from occipital centers. Although visual observation of the eye movements of brain damaged and normal *Ss* has shown that the former exhibit markedly more frequent and grosser eye movements during "fixation," this finding has yet to be confirmed by ophthalmographic recording.

In view of the wide variations in the locus, extent, and nature of brain damage there seems little justification for attempting to establish a relationship between a sensory aftereffect and brain damage. It is possible that this simple test may prove useful in differentiating between types of damage rather than differentiating it from other conditions. If the hypothesis outlined here is supported by subsequent data, a more efficient technique would be that of observing the magnitude and direction of eye movement during the period when *S* is attempting to fixate. Such a method would have the advantage of dispensing with subjective reports of an experience which may be an indirect outcome of brain damage.

Summary

Recent studies have shown that with brain damaged *Ss* the aftereffect of seen movement from a rotating spiral pattern is either absent or weaker than that experienced by normal *Ss*. Some studies have failed to confirm this observation.

A necessary condition for the occurrence of this aftereffect is steady fixation of a stationary point in or near the moving stimulus pattern so that the retinal image of the pattern moves relative to the retina. Voluntary fixation and eye movement is affected in certain cases of brain damage so that individual's eyes tend to follow the moving pattern. It is suggested that the weakness or absence of the aftereffect from a moving pattern derives from this failure of voluntary fixation.

Since the main centers controlling voluntary eye movement and fixation are located in the frontal region of the brain, it would be expected that damage affecting that region would have the most marked effect on the aftereffect.

Inconsistencies between the results from various studies probably derive in large part from a failure to consider the locus, extent, and nature of brain damage as well as the inadequacies of the methods used to measure the aftereffect.

Some preliminary results in support of the eye movement hypothesis have been presented.

REFERENCES

Aaronson, B. S. Age intelligence, aphasia and the spiral aftereffect in an epileptic population. *J. clin. Psychol.*, 1958, **14**, 18–21.

Davids, A., Goldberg, L., & Laufer, M. W. The relation of the Archimedes spiral after-effect and the Trail Making Test to brain damage in children. *J. consult. Psychol.*, 1957, **21**, 429–433.

Day, R. H. On interocular transfer and the central origin of visual after-effect. *Amer. J. Psychol.*, 1958, **71**, 784–789.

Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–797.

Gallese, A. J. Spiral after-effect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.

Garrett, E. S., Price, A. C., & Deabler, H. L. Diagnostic testing for cortical brain impairment. *Arch. Neurol. Psychiat.*, 1957, **77**, 223–225.

Goldberg, L. R., & Smith, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, **22**, 153–157.

Holland, H. C., & Beech, H. R. The spiral aftereffect as a test of brain damage. *J. ment. Sci.*, 1958, **104**, 466–471.

Holmes, G. The cerebral integration of ocular movements. *Brit. med. J.*, 1938, **2**, 107–112.

Johnson, D. E., Bauer, R. W., & Brown, D. R. Visual afterphenomena in diagnosis. *J. consult. Psychol.*, 1959, **23**, 90.

Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.

Penfield, W., & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 1937, **60**, 389–443.

PENFIELD, W., & RASMUSSEN, T. *The cerebral cortex of man*. New York: Macmillan, 1950.

PHILBRICK, E. B. The validity of the spiral after-effect as a clinical tool for diagnosis of brain pathology. *J. consult. Psychol.*, 1959, **23**, 39–43.

PICKERSGILL, M. J., & JEEVES, M. A. Aftereffect of movement produced by rotating spiral. *Nature*, 1958, **182**, 1820.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

SAUCER, R. T. Nature of perceptual processes. *Science*, 1953, **117**, 556–558.

SAUCER, R. T. Processes of motion perception. *Science*, 1954, **120**, 806.

SHERRINGTON, C. S. *Man on his nature*. Cambridge: Cambridge Univer. Press, 1940.

SPIVACK, G., & LEVINE, M. The spiral after-effect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, **25**, 1–11.

WALSH, E. G. *Physiology of the nervous system*. London: Longmans Green, 1957.

WOHLGEMUTH, A. On the after-effect of seen movement. *Brit. J. Psychol. Monogr. Suppl.*, 1911, No. I.

(Received July 29, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 316–323

PSYCHOTHERAPISTS' CLINICAL JUDGMENTS AND ATTITUDES TOWARDS PATIENTS[1]

MARTIN S. WALLACH AND HANS H. STRUPP
School of Medicine, University of North Carolina

Theoretically, and from the standpoint of research, the therapist appears to be finding his way back into therapy. This trend is in keeping with a renewed interest in countertransference or, more broadly, the therapist's contribution to the treatment process. Mensh (1956) states, ". . . study of the therapist is now more explicitly indicated, an emphasis which has appeared with trends toward introducing experimental design qualities into research in psychotherapeutic process" (p. 342). Butler (1952) asserts that ". . . exact observation of the therapist is a necessary precondition to understanding the behavior of the client" (p. 378); similarly, Gitelson (1952), Jackson (1956), and Wolberg (1954) stress the importance of therapist variables. Perhaps the primary impetus for focusing upon the therapist stems from a belief that orthodoxy is waning and "Viewing therapy as a very intimate relationship and recognizing a trend toward less restrictive therapist roles, the expectation arises that differential therapist characteristics would have systematic effects on process and outcome" (Winder, 1957, p. 316).

Previous attempts to explore the relationships between therapist variables and therapist performance in treatment have considered mainly those factors which most readily lend themselves to objective quantification: i.e., the therapist's professional affiliation, theoretical allegiance, length of experience, number of hours of personal analysis, etc. The potential importance of less tangible personality variables, long recognized in Freud's conception of countertransference, is beginning to attract the attention of investigators seeking objective documentation. For instance, Parloff (1956) found that the therapist relates most satisfactorily to those patients who most clearly approximate his "Ideal Patient." Strupp (1958) has shown that the therapist's attitude toward the patient was significantly correlated with the form of therapy, frequency and length of treatment, the kinds of technical problems expected in therapy, and certain qualitative aspects of the techniques the therapist expected to employ. These observations strongly support the notion that emotional and attitudinal factors in the therapist may have an important bearing upon his clinical evaluations.

The therapist, as a function of his life experiences, approaches each initial interview with needs, expectations, and wishes of his own. It may be hypothesized that, if his expectations are sufficiently realized, he will consider the situation as "rewarding," and a "warm" attitude toward the patient is likely to develop. More specifically, if in an initial interview the prospective patient approximates the therapist's conception of an "ideal patient," he may develop a warm attitude toward the patient. While the conception of "ideal patient" may differ from therapist to therapist, in a general sense, it may be assumed to imply a certain congruence between the kind of help the patient is seeking and the kind of help the therapist is able to provide, as well as an explicit or implicit belief on the patient's part that the treatment will be effective. Patients exhibiting such hopes

[1] This paper is a summary of a dissertation submitted in partial fulfillment of the PhD degree to the Department of Psychology, University of North Carolina.

This research is part of a larger project supported by a research grant (M-2171) from the National Institute of Mental Health, Public Health Service.

and beliefs may be said to be "motivated for therapy" or "ready for a psychotherapeutic experience." Thus, patients whose beliefs, desires, and expectations are similar—in some respects at least—to corresponding expectations held by the therapist may be said to arouse a warmer attitude in the therapist. This attitude, it is hypothesized, will be sufficiently strong to give rise to a kind of "halo effect" overshadowing certain "reality factors." Operationally, this may be evidenced by the character of the therapist's perception of the patient, treatment planning, and the like.

The initial halo effect, it is further hypothesized, will be self-realizing. This is to say, having achieved a "comfortable" or "warm" feeling toward the patient, the therapist will continue to focus on perceptions which are likely to engender further this effect, and he will propose a therapeutic regime which will reflect this attitude. Moreover, he may be more inclined to view the patient as a "good patient" and work toward a closer therapeutic contact with the patient. The present investigation aims to explore the relationships between the therapist's attitude toward the patient on the one hand and his perceptions of the patient and treatment plans on the other. The causal direction, of course, remains an open question.

Hypotheses

1. A patient's verbal or nonverbal expectations concerning his "motivation for therapy" or "readiness for therapy" is an important determinant of the therapist's attitude toward the patient. In particular, the more highly motivated patient will engender in the therapist a warmer attitude.

2. The therapist's attitude toward the patient, whether a function of the patient's motivation for therapy or other factors, is correlated with the therapist's perception of, and treatment plans for, the patient. More specifically, a warmer therapist attitude toward the patient is associated with more favorable perceptions of the patient, including clinical judgments, prognostic estimates, and treatment plans implying closer therapeutic contact.

Procedure

To test these hypotheses it was planned to present a group of therapists with a patient manifesting relatively high motivation for therapy, and a comparable group of therapists with the same patient showing relatively low motivation for therapy. Furthermore, to lend greater generalizability to the findings, therapists were presented with two patients.

Two written case histories of neurotic patients were adapted for experimental purposes. Each case history consisted of alternate forms which were identical in all respects, except that one form presented the patient as highly motivated and the second form as poorly motivated for therapy. "Motivation" was communicated primarily in terms of the patient's eagerness to seek and accept psychotherapeutic help. Each case history was followed by an identical questionnaire of 27 items eliciting the respondents' clinical impressions, treatment plans, and attitudes toward the patient. Subjects were presented with the case histories of both patients, the degree of motivation being varied systematically. In addition, several biographical questions were appended to the second questionnaire.

Sample

Data were collected from medical psychotherapists actively engaged in the practice of psychotherapy. Three sources were utilized: (*a*) physicians known to the investigators personally, (*b*) colleagues of psychologists at selected installations, and (*c*) psychiatrists selected at random from the Directory of the American Psychiatric Association. Of

TABLE 1

Biographical Data of Respondents

	N	Male	Female	Mean Age	Mean Hours of Personal Analysis	Mean Years of Experience
Psychiatric Residents	11	10	1	33.91	140 (2)[a]	3.05
Psychiatrists	37	34	3	35.86	487 (26)	7.27
Psychoanalysts	34	30	4	44.47	625 (34)	14.31
Totals	82	74	8	39.17	549 (62)	9.62

[a] Figures in parentheses indicate number of respondents who had personal analysis.

300 sets of material distributed personally or by mail, 82 were included in the analysis. One hundred and eighty-four sets were never returned, 26 respondents asked to be excused, and 8 respondents completed and returned the questionnaires after the analysis of the data was well in progress. Analysis of the biographical data (age, sex, professional level, years of experience, number of hours of personal analysis) supplied by respondents indicated that the sample was drawn from a relatively homogeneous universe of experienced medical psychotherapists (see Table 1), and that in terms of these variables the various subsamples showed no significant variations.

Results

Position Effects

As a first step in the statistical analysis it seemed important to investigate the possible influence of order of presentation (patient and degree of motivation) upon questionnaire responses. Since 24 multiple-choice items were answered under four experimental conditions, 96 comparisons were called for. Chi squares and *t* tests of these comparisons yielded 15 results significant at beyond the .1 level. These significant position effects appeared to be largely due to the fact that one of the patients seemed objectively more disturbed, so that respondents studying the male (more disturbed) patient first tended to attribute greater personality strengths and assets to the female (less disturbed) patient when her case was presented second.

Effects of Communicated Patient Motivation

Chi square or *t* tests were employed to ascertain effects of the patient's motivation, as communicated in the case histories, upon responses to the 24 multiple-choice items. Five figures significant at or beyond the .05 level were obtained for the male case, and nine figures significant at this level were obtained for the female case.

When Mr. J. was presented as possessing high motivation for therapy, respondents tended to attribute to him greater ego strength, more anxiety, more insight into his problems, greater motivation for therapy (check item), and they expressed greater willingness to accept the patient for treatment. Contrary to the hypothesis, therapists' attitudes toward the patient did not vary significantly. However, there was a (statistically nonsignificant) trend in the predicted direction—a slight tendency for therapists to claim a warmer attitude toward the patient when he was presented as being more highly motivated for therapy.

When Mrs. D. was presented as possessing high motivation for therapy, respondents estimated ego strength, insight into her problems, and motivation (check item) as being greater. They likewise judged her social adjustment as more adequate, more often recommended psychotherapy or psychoanalysis without adjunctive forms of treatment (drugs, shock), and indicated that they would strive to accomplish a more extensive change in the patient's character structure. Presenting Mrs. D. with high motivation also accounted for a more favorable prognosis with treatment and a greater willingness on the part of the therapist to take her into treatment. Finally, as hypothesized, therapists claimed a warmer attitude toward the patient when high motivation was communicated.

Effects of Perceived Patient Motivation

The preceding comparisons were made in terms of the level of motivation communicated in the case material which did not necessarily coincide with the patient's motivation as actually perceived by the respondents. For the male case, there was high agreement between the kind of motivation communicated and the kind of motivation perceived by respondents. For the female case, however, the disparity between communicated and perceived motivation was greater, as evidenced by 13 shifts (16%). Ten of these respondents "misread" the case history intended to convey low motivation by estimating the patient's motivation to be relatively high. Consequently, it was thought advisable to determine the relationship between perceived motivation and other variables, and chi square tests were performed between the respondent's estimate of the patient's motivation and each of the other 23 multiple-choice items.

For the male case, questionnaire items already cited as being significantly related to communicated motivation were again significantly related to perceived motivation. In addition, five other items were significantly

related to perceived motivation at or beyond the .05 level. Respondents who perceived Mr. J. to be highly motivated also tended to estimate social adjustment to be more adequate, suggested more frequent therapeutic sessions, stated that they would be less willing to alter the kind of therapy they usually conduct, estimated the prognosis to be more favorable with treatment, and found it easier to empathize with the patient.

For the female case, estimates of greater ego strength, more insight, better social adjustment, more favorable prognosis with treatment, greater willingness to accept the patient in treatment, and warmer therapist attitudes accompanied perceived motivation, as they did communicated motivation. On the other hand, perceived high motivation (unlike communicated high motivation) was not significantly related to form of treatment or to the extensiveness of attempted change in the patient's character structure. However, respondents who perceived Mrs. D. as highly motivated for therapy also tended to judge her as being more emotionally mature and less seriously disturbed.

Effects of Therapists' Attitudes

In order to test the second hypothesis, the respondents' stated attitudes toward the patient were compared, by chi square, with the 23 remaining multiple-choice items.

For the male case, eight items varied significantly with the therapist's attitude toward the patient, and only one of these (willingness of the therapist to alter his usual way of conducting therapy) was in the opposite of the predicted direction. Briefly, a positive thera-

TABLE 2

SUMMARY OF STATISTICAL COMPARISONS BETWEEN PATIENTS' MOTIVATION FOR THERAPY, THERAPISTS' ATTITUDE, AND OTHER QUESTIONNAIRE ITEMS

($N = 82$)

	Male Case		Female Case	
Therapists' Estimates	Motivation[b]	Therapist Attitude	Motivation[b]	Therapist Attitude
Diagnosis	*NS*[a]	.02	*NS*	*NS*
Ego strength	.001	.01	.05	*NS*
Anxiety	.05	*NS*	*NS*	*NS*
Insight	.02	.001	.01	.05
Emotional maturity	*NS*	*NS*	.01	*NS*
Social adjustment	.05	*NS*	.01	.05
Degree of disturbance	*NS*	*NS*	.01	*NS*
Similarity to patients in therapy	*NS*	*NS*	*NS*	*NS*
Kind of treatment	*NS*	*NS*	*NS*	*NS*
Motivation		*NS*		.001
Chances of acting out	.1	*NS*	*NS*	*NS*
Frequency of sessions	.05	*NS*	*NS*	*NS*
Extensiveness of change in character structure	*NS*	*NS*	.1	*NS*
Length of treatment	*NS*	*NS*	*NS*	*NS*
Permissiveness	*NS*	*NS*	*NS*	*NS*
Encourage free association	*NS*	*NS*	*NS*	*NS*
Recommendations to patient	*NS*	*NS*	*NS*	*NS*
Change of usual therapy	.01	.05	*NS*	*NS*
Prognosis without therapy	*NS*	.05	*NS*	*NS*
Prognosis with therapy	.05	.02	.001	.05
Willingness to accept patient in Rx.	.01	.001	.001	.001
Ease of empathizing with patient	.01	.001	*NS*	.001
Environmental stress	*NS*	*NS*	.1	*NS*
Therapist's attitude	*NS*		.001	

[a] *NS*—not significant at .10 level (chi square, two-tailed tests).
[b] Patient's motivation as perceived by therapist.

pist attitude was found to be associated with: a diagnosis of psychoneurosis rather than one of psychosis or personality disorder, estimates of greater ego strength and insight, *less* willingness to alter his usual therapeutic procedure, more favorable prognostic estimates with or without treatment, greater willingness to accept the patient for treatment, and greater ease of empathizing with the patient.

For the female case, six items showed a significant relationship with attitude at the .05 level, and all differences were in the predicted direction. Therapists who claimed a more positive attitude toward Mrs. D. considered her to possess greater insight into her problems, more adequate social adjustment, and higher motivation for therapy. Warmer attitudes toward this patient were also accompanied by estimates of a more favorable prognosis with treatment and greater ease of empathizing with the patient.

In addition to the statistically significant results already reported, more than 75% of the nonsignificant comparisons showed trends in the predicted direction. The figures for these comparisons are given in Table 2.

Additional Analyses

The theoretical possibility that basic dimensions might underlie the observed relationships among questionnaire items led to

TABLE 3

INTERCORRELATION AMONG SELECTED ITEMS

(N = 75 to 82)

Item	Ego Strength	Insight	Social Adjustment	Motivation	Alter Rx.	Accept Patient	Prognosis	Empathy
				Male Case				
Insight	.30							
Social adjustment	.61	.28						
Motivation	.32	.50	.27					
Change of usual therapy	−.30	−.31	−.06	−.34				
Prognosis with therapy	.46	.32	.32	.26	−.51			
Willingness to accept patient	.34	.30	.20	.40	−.30	.40		
Ease of empathizing with patient	.43	.14	.25	.22	−.02	.22	.28	
Therapist's attitude	.52	.23	.10	.22	−.29	.42	.55	.44
				Female Case				
Insight	.49							
Social adjustment	.37	.35						
Motivation	.40	.65	.37					
Change of usual therapy	−.27	−.15	−.07	−.18				
Prognosis with therapy	.41	.37	.28	.55	−.52			
Willingness to accept patient	.19	.39	−.08	.46	−.21	.68		
Ease of empathizing with patient	.08	.35	.20	.25	−.01	.32	.40	
Therapist's attitude	.30	.40	.31	.59	−.07	.48	.54	.44

an attempt to isolate such factors. Since a factor analysis employing the intercorrelation of all 27 items seemed impractical, only the 9 items already found to yield significant comparisons were utilized. For each case, these 9 items were intercorrelated (product-moment coefficients of correlation). A centroid factor analysis was performed for each patient separately, three factors were extracted in each analysis, and both factor analyses were then rotated to an oblique simple structure.

The similarity of patterns of the intercorrelation matrices (see Table 3) for each case suggests that certain items are consistently related to other items and that these relationships are rather stable. That is, patient differences as represented by the two cases in this investigation have little effect upon the degree to which various therapist estimates and judgments are correlated, and the observed association between items can be assumed to be mainly a function of therapist variables.

The three factors extracted for each patient are difficult to interpret, and the findings for the two patients often failed to parallel each other. However, common to both factor analyses was the emergence of what might be termed a "good patient" factor. For Mr. J. this factor includes high loadings on items which give the following description: the therapist is willing to accept a well motivated, insightful patient for his usual treatment procedure. For Mrs. D. high loadings occur on such items as: the therapist perceives the patient as having insight, high motivation for therapy, and is easy to empathize with; the therapist has a warm attitude toward the patient and is eager to accept her for treatment.

Discussion

Since this investigation is not based on observations of psychotherapists in their natural setting, questions may be raised about the correlation between their behavior responding to questionnaires and their behavior in actual psychotherapy when they have the opportunity for direct acquaintance with the patient. As one of the authors (Strupp, 1958) pointed out in a somewhat different context,

> It is not contended that the behavior of the psychiatrists in an initial interview with a "real" patient bears a one-to-one relationship to his performance in the experiment. . . . What is asserted is that the evaluative processes, as they were studied under experimental conditions, are sufficiently similar to the "normal" behavior of the therapist to generate important hypotheses about his behavior in a first interview. Hypotheses thus derived are valuable if they raise further questions for research, and if they can be applied to practice and training in psychotherapy (p. 64).

Asch (1952) comes to a similar conclusion in evaluating the generalizability of his experimental findings which were based on descriptions rather than on observation of actual persons: "certain processes occurring in response to indirect data about persons not present are also at work in reaction to actual persons" (p. 219). Nevertheless, the possibility that concrete relations with persons introduce additional variables cannot be dismissed lightly.

Regarding our findings, it is seen that certain therapist estimates pertaining to perceptions of, and treatment plans for, the patient are related to the therapist's professed attitude toward the patient. Estimates of diagnosis, ego strength, inclination to modify therapy, prognosis without treatment, social adjustment, and motivation for therapy appear to vary with the therapist's attitude and the patient's degree of disturbance (it will be recalled that the female patient was less seriously disturbed than the male patient).

It should be pointed out that only 10 of the 24 multiple-choice items yielded results which were consistently statistically significant. At the same time, the fact that these results were obtained for two patients adds weight to the evidence. Furthermore, the fact that the vast majority of statistically nonsignificant comparisons were in the predicted direction, coupled with the respondents' frequently expressed claim that the written case histories did not allow them to use all the cues they ordinarily employ, suggests that many of the relationships are probably underestimates.

The direction of causation obviously remains indeterminate. Thus, it is not clear whether the therapist's attitude "causes" the character of his clinical estimates, whether

the therapist feels more favorably disposed toward patients who are better therapeutic risks to start with, or whether both possibilities are more complex functions of other variables. It seems clear, however, that insight, or the capacity for insight, is a key variable entering into prognostic estimates, the therapist's eagerness in accepting the patient for treatment, and his (self-rated) ability to empathize with the patient. Therapists tend to profess a more positive attitude towards a patient who meets these criteria. On the other hand, the data suggest that the patient's motivation for therapy influences the therapist's attitude in the predicted direction only when the degree of disturbance is not too great. In other words, a therapist "likes" a patient who has the capacity for insight and is eager to be helped provided the clinical evidence supplies no contraindications.

These findings show some interesting parallels to the ingenious experiments of Asch (1946) on the formation of impressions of personality. Among many other findings, Asch reports that "a central trait determines the content and the functional place of peripheral traits within the entire impression" but that central characteristics are "themselves affected by the surrounding characteristics" (p. 276). In his studies, the "warm-cold" dimension appeared to occupy such a central position. In this investigation, motivation for therapy similarly seems to determine a number of clinical estimates as well as the therapist's attitude. (For further evidence on this point, see the study by Strupp & Williams, 1960). Our results also seem to corroborate Asch's conclusions that one can see another person only as a unit, that qualities enter into a particular relation and become organized, and that the system of the traits points to a necessary center. Altogether it seems that application of Asch's technique to the investigation of the formation of clinical judgments presents a fascinating research task.

Regardless of patient's motivation for therapy, however, therapists profess a more positive attitude towards patients who are less severely disturbed. While this investigation has nothing to say about therapeutic outcomes, it dovetails with the findings of a number of such studies summarized by Luborsky (1959): "Those who stay in treatment improve; those who improve are better off to begin with than those who do not; and one can predict response to treatment by how well they are to begin with" (p. 324). We may add, on the basis of our findings, that therapists are more enthusiastic about treating patients whom they view as having a better prognosis—perhaps, because they are perceived as having greater capacity for profiting from psychotherapeutic procedures which rely both on the patient's intellectual as well as emotional understanding. Our first hypothesis may thus be recast: A "warm" therapist attitude toward a patient is more likely to be present when the patient is well motivated for therapy, provided the severity of his emotional disturbance is not too great.

The one result which ran counter to the prediction showed a negative relationship between a favorable therapist attitude and his willingness to alter his usual method of therapy. In the context of other findings here reported this disinclination to alter his usual method of therapy seems to reflect the therapist's conviction that the patient can profit from the kind of psychotherapy the therapist employs most often and has the greatest faith in and that there is no need to press into service other techniques and therapeutic adjuncts—the use of which may, at times, signify a subtle rejection of the patient as a person.

Estimates of the patient's insight into his problems showed the highest degree of association with the patient's motivation for therapy and the therapist's attitude. This finding supports the notion that estimates of insightfulness are highly descriptive of the good (potential) learner as far as psychotherapeutic achievement is concerned. It also substantiates the belief that psychotherapeutic success requires, among other things, the patient's ability to achieve insight and the therapist's ability to foster a warm, understanding, and accepting environment. Fiedler (1950) found these factors to be mentioned prominently in therapists' descriptions of the ideal therapeutic relationship, and he observed that agreement on these variables was

higher among experienced therapists regardless of theoretical orientation than among experienced and inexperienced therapists of like orientation.

The demonstrated relationships between therapists' attitudes and their clinical evaluations and treatment plans in general support the second hypothesis. This may also be interpreted as discrediting the conception of an impersonal, mirror-like therapist, a view which has already become outmoded. By drawing attention to these associations, this study supports the increasing emphasis given to the effects of personality variables in the therapist upon his clinical operations. The extent to which these variables play a part in the therapist's intricate work remains to be explored much more exhaustively, but it is clear that they must come to occupy a more central place in selection, training, and therapeutic practice. From a methodological point of view, this study demonstrates the feasibility and potential value of quasiclinical experimental designs, which permit the introduction of controls not otherwise possible.

Summary

This investigation deals with an exploration of two hypotheses positing an association between therapists' attitudes toward patients and clinical evaluations and treatment plans as a function of the patient's motivation for therapy. Eighty-two respondents, all of whom were practicing medical psychotherapists, read two case histories of neurotic patients. Alternate forms of the case histories were employed to vary systematically the patient's motivation for therapy. Respondents recorded their clinical impressions, estimates, and attitudes by means of a 27-item questionnaire which followed each case presentation.

The hypothesized association between therapists' attitudes toward the patient and their perceptions of, and treatment plans for, the patient appear in general to be substantiated. It is pointed out that the direction of causation remains indeterminate and that the degree of relationship between the therapist's "usual" performance and his performance in the present experimental situation is unknown. Nevertheless the evidence is believed to support the assertion that personality factors of the therapist are an integral part of his clinical judgments and therapeutic procedures. A better understanding of these relationships might make an important contribution to more effective clinical practice as well as aid in the selection of candidates for training in psychotherapy.

REFERENCES

Asch, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

Asch, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

Butler, J. M. The interaction of client and therapist. *J. abnorm. soc. Psychol.*, 1952, **47**, 366–378.

Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.

Gitelson, M. The emotional position of the analyst in the psychoanalytic situation. *Int. J. Psycho-Anal.*, 1952, **33**, 1–10.

Jackson, D. D. Countertransference and psychotherapy. In Frieda Fromm-Reichmann & J. L. Moreno (Eds.), *Progress in psychotherapy—1956*. New York & London: Grune & Stratton, 1956. Pp. 234–238.

Luborsky, L. Psychotherapy. In P. R. Farnsworth & Q. McNemar (Eds.), *Annu. Rev. Psychol.* Vol. 10. Palo Alto, Calif.: Annual Reviews, 1959. Pp. 317–344.

Mensh, I. N. Research in counseling and psychotherapeutic process. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. Vol. 2. New York & London: Grune & Stratton, 1956. Pp. 340–360.

Parloff, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, **52**, 5–10.

Strupp, H. H. The psychotherapist's contribution to the treatment process. *Behav. Sci.*, 1958, **3**, 34–67.

Strupp, H. H., & Williams, Joan V. Some determinants of clinical evaluations of different psychiatrists. *AMA Arch. gen. Psychiat.*, 1960, **2**, 434–440.

Winder, C. L. Psychotherapy. In P. R. Farnsworth & Q. McNemar (Eds.), *Annu. Rev. Psychol.* Vol. 8. Palo Alto, Calif.: Annual Reviews, 1957, Pp. 309–330.

Wolberg, L. R. *The technique of psychotherapy*. New York: Grune & Stratton, 1954.

(Received August 3, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 324–327

OPERANT CONDITIONING OF HOSTILE VERBS IN RELATION TO EXPERIMENTER AND SUBJECT CHARACTERISTICS [1]

DONALD C. FERGUSON AND ARNOLD H. BUSS
University of Pittsburgh

Virtually all operant verbal conditioning studies have used neutral materials. One of the exceptions is an experiment by Buss and Durkee (1958), who used a modification of Taffel's procedure (1955). A list of past tense verbs was compiled, and judges sorted them into three categories: neutral, mildly hostile, and intensely hostile. Three verbs, one from each category, were placed on a card, and *Ss* were instructed to make up a sentence using one of the verbs. There were 60 such cards (trials). One group of *Ss* was reinforced for using intensely hostile verbs, and another group was reinforced for using neutral words. It was found that intensely hostile words were learned at a faster rate than neutral words. There were also sex differences in frequency of using intensely hostile words.

This study was followed up by Binder, McConnell, and Sjoholm (1957), who used only the neutral and mildly hostile verbs. There were two groups of *Ss*, both reinforced for using hostile verbs. One group was run by a large, aggressive man and the other by a diminutive, mild, nonassertive woman. The group run by the "neutral" woman learned significantly better than the group run by the aggressive man.

Evidently characteristics of the *E* affect learning of hostile material. However, it is not clear from the Binder et al. (1957) study which characteristics are important. One *E* was large, male, and aggressive; the other *E* was small, female, and neutral. It is doubtful that size of *E* is a major variable, which leaves aggressiveness and sex. Which of these variables retarded learning, aggressiveness or masculinity? Perhaps both affected learning. The Binder et al. (1957) experiment does not furnish an answer because aggressiveness and gender of the *E* were confounded. The present study investigated these variables separately, i.e., a male and a female *E* each played an aggressive or a neutral role with *Ss*. In addition, sex of *Ss* was a variable in an attempt to discover interactions between characteristics of the *E* and gender of the *S* in the operant verbal conditioning of hostile material.

[1] This report is based on a master's thesis by the senior author under the direction of the junior author.

METHOD

Experimental Design

There were four groups of *Ss*. Two groups were run by a man and two groups by a woman. The male *E* was the senior author, and the female *E* was his wife.[2] Each *E* played a neutral role with one group and an aggressive role with the other. Preexperimental rehearsal helped to make enactment of each role similar for the two *E*s.

The neutral *E* was patient, calm, and courteous in an attempt to establish a nonhostile experimental climate. The aggressive *E* was brusque and unfriendly. The *S* was made to wait if early or on time and criticized if late. The *E* was impatient and tended to scowl and sneer. Reports from *Ss* at the termination of the experiment revealed that the two roles (neutral or aggressive) were enacted appropriately.

Procedure

The procedure was the same as that of Binder et al. (1957). The *S* was presented with a sample card on which was typed the verbs WALKED and RAN, and the pronouns HE, THEY, SHE. He was instructed to make up a sentence using one of the verbs and one of the pronouns. Then he was presented with a series of 140 similar cards, each card constituting a trial. On these cards one verb was neutral and one was mildly hostile.

[2] The writers wish to thank Elsie Ferguson for serving as an *E*.

The first 20 trials were used to establish an operant level of responding. Starting with Trial 21 the *E* said "Good" each time a hostile verb was used. This procedure was the same for all *S*s.

Subjects

The *S*s were 80 college students recruited from psychology classes. The *S*s were randomly assigned to the four experimental groups, with the restriction that half the *S*s in each group were men and half were women.

Results

An examination of the learning curves revealed no systematic sex differences. Therefore the learning curves of male and female *S*s were combined. The data bearing on aggressiveness of the *E* are presented in Fig. 1. The neutral *E* curve is a composite of the data from the male and female *E*s, and the aggressive *E* curve combines groups run by the male and the female *E*s. The aggressive *E* curve is smoother and has a steeper slope than the neutral *E* curve. The superiority of learning with a neutral *E* is especially evident during the last 60 trials (Blocks 5, 6, and 7).

The data bearing on gender of the *E*, presented in Fig. 2, indicate that the sex of the *E* has little effect on learning. The female *E* curve starts at a slightly higher level than the male *E* curve. However, this difference disappears during the second block of trials, and thereafter the curves are similar.

None of the learning curves presented in the two figures is smooth, but the irregularities do not fall into any pattern. In this respect they resemble the curves obtained in most operant verbal conditioning studies, in which uncontrolled variables appear to be affecting the conditioning process. In this context neither the shape of the curve (which reflects random irregularities) nor the overall height of the curve is of particular importance. Overall height is determined not only by increments in learning but also by initial level of responding. Therefore, the crucial measure of conditioning is the increment in learning, i.e., the net increase in frequency of hostile verbs above the operant level of responding. The score used was: frequency for last 20 trials minus frequency for the first (operant) 20 trials.[3]

An analysis of variance of these scores is presented in Table 1. The only significant *F* is for aggressiveness of the *E*. Sex of the *E*, sex of the *S*, and interactions among the variables were not significant. This analysis of

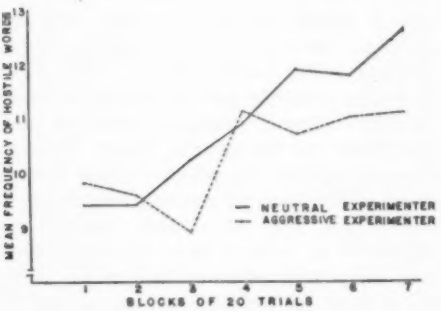

Fig. 1. Conditioning of hostile verbs with a neutral or aggressive *E*.

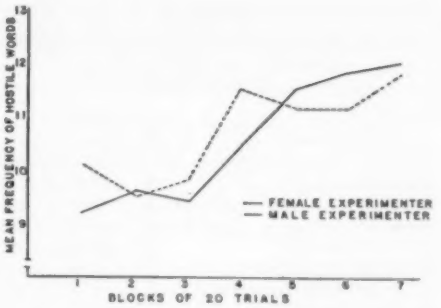

Fig. 2. Conditioning of hostile verbs with a male or female *E*.

[3] One of the writers (Buss) has conducted several experiments on operant verbal conditioning with both neutral and hostile materials. These studies have yielded learning curves that vary not only from one *S* to the next but from one sample to the next. The lack of smoothness in curves has not obliterated the presence of group differences, probably because such differences were large, but it warns against placing much faith in intertrial variations. An analysis of trend, such as that used by Binder et al. (1957), takes advantage of variations among blocks of trials. Since such intertrial variations tend to be unreliable, they probably consist mainly of error variance. Therefore, trend analysis was not used in testing the significance of the present data. Rather, an analysis of variance of the net increase in response frequency was employed. The net increase measure seems best because it is the most conservative and the most stringent measure of learning, since it does not take advantage of random variations within blocks of trials.

TABLE 1

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR FREQUENCY OF HOSTILE VERBS

Source	*df*	Mean Square	*F*
Aggressiveness of *E*	1	66.61	4.24*
Sex of *E*	1	30.01	1.91
Sex of *S*	1	1.01	.06
Sex of *E*×Sex of *S*	1	10.51	.67
Sex of *E*×Aggressiveness of *E*	1	2.11	.13
Sex of *S*×Aggressiveness of *E*	1	2.11	.13
Sex of *S*×Sex of *E* ×Aggressiveness of *E*	1	.32	.02
Within groups	72	15.70	
Total	79		

* $p < .05$.

variance confirms the trends noted in the learning curves. Only aggressiveness of the *E* significantly affects learning, an aggressive *E* tending to retard conditioning.

Since the curves in Figs. 1 and 2 differ in overall height, another score was used: total frequency of hostile verbs. An analysis of variance of these scores yielded no significant *F*. Evidently none of the variables influenced total frequency of hostile verbs.

DISCUSSION

Binder et al. (1957) found that an aggressive male *E* retarded learning in comparison to a neutral female *E*. The results of the present study indicate that the effective variable is not the sex but the aggressiveness of the *E*. Operant conditioning of hostile verbs is slower with an aggressive *E* than with a neutral *E*. In attempting to explain this finding two possibilities should be considered. First, when the *E* was aggressive, he may have been a less potent reinforcing agent. His saying "Good" would be less reinforcing than the neutral *E*'s "Good" and would lead to slower learning.

An alternative explanation is that *E*'s being aggressive angered the *S*. The *S* could not aggress openly against the *E*, because of his higher status as an *E*. Therefore, it would be necessary to repress or suppress aggressive tendencies, and such inhibition would generalize to the hostile verbs. The *S* would then tend to inhibit verbalizing the hostile choice, and this would retard conditioning. Which of these two explanations is correct will be decided by studies in which other kinds of material are conditioned and the *E* assumes nonneutral roles other than aggressiveness.

Two of the groups in the present study may be compared with their counterparts in the Binder et al. (1957) study: the group run by the neutral female *E* and the group run by aggressive male *E*. The neutral female *E* curves in the two experiments are similar, with the curves of the present study starting slightly higher and ending slightly higher; the slopes appear almost identical. The aggressive male *E* curves start from the same operant level of responding, but the curve of the present study has a steeper slope, ending at a higher point. Considering the variability in these curves from one block of trials to the next, it is doubtful that these differences in the aggressive *E* curves are significant.

In both studies the slope of the neutral *E* curve was significantly steeper than that of the aggressive *E* curve, i.e., the increment in frequency of hostile verbs was significantly greater for the *S*s run by a neutral *E*. Concerning the total frequency of hostile verbs, the two experiments differ. Binder et al. (1957) found a significant difference between the neutral *E* and aggressive *E* groups in total frequency of hostile verbs, whereas the present investigators did not, although the difference was in the same direction. However, as noted earlier total frequency is influenced not only by increments in learning but also by initial level of responding and by irregularities in the learning curves. Examination of Fig. 1 reveals that the curve of the aggressive *E* started higher and because of irregularities in the curve, it overlapped the neutral *E* curve at the fourth block of trials. These two facts probably account for the difference in total frequency of hostile verbs not being significant in the present study. However, since the crucial dependent variable is *increment* in frequency of response, not height of the curves, this discrepancy between the Binder et al. (1957) study and the present one seems unimportant. Such minor differences must be expected so long as we are dealing with curves that are subject to random irregularities. With respect to the major

finding—that an aggressive *E* retards conditioning of hostile verbs—the two studies are in agreement.

Summary

This experiment investigated operant conditioning of hostile verbs in relation to aggressiveness of the *E*, sex of the *E*, and sex of the *S*. The *S*s were presented with two verbs and three pronouns on a card. One verb was hostile and the other neutral. The *S*s made up a sentence using one verb and one pronoun. They were reinforced for using hostile verbs. It was found that only the aggressiveness of the *E* significantly affected conditioning: an aggressive *E* retards learning in comparison to a neutral *E*.

REFERENCES

Binder, A., McConnell, D., & Sjoholm, Nancy A. Verbal conditioning as a function of experimenter characteristics. *J. abnorm. soc. Psychol.*, 1957, **55**, 309–314.

Buss, A. H., & Durkee, Ann. Verbal conditioning of hostile and neutral verbs. *J. consult. Psychol.*, 1958, **22**, 415–418.

Taffel, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

(Received August 9, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 328–332

THE EFFECT OF A STRUCTURED INQUIRY ON RORSCHACH SCORES [1]

MELVIN ZAX AND GEORGE STRICKER
University of Rochester

In spite of the fact that the inquiry is an integral part of the Rorschach examination there is little unanimity among experts concerning the proper way to conduct it, nor is much empirical data available on the effects of various procedures. All prominent Rorschach theorists agree that the examiner should adopt a nonleading approach when conducting the inquiry. When they operationalize their recommendations, however, it is apparent that they disagree on what constitutes "leading the subject."

Klopfer and Kelley (1946) advocated procedures such as offering alternative concepts or comparing a previously given percept to a later one in order to establish the presence or absence of a specific determinant. In probing for movement they also sanctioned questioning about the positions of the figures seen. As a last resort in ruling on the use of color they agreed with Beck (1950) in employing such questions as, "Suppose this were exactly the same shape as now, but gray or black, would you still think it is . . . ?"

Klopfer, Ainsworth, Klopfer, and Holt (1954) cautioned against suggesting any particular determinant to subject (*S*) and, as a way of avoiding this, advised sticking closely to the terms and language employed by *S* himself. They were also critical of the practice of comparing the Rorschach plate with the small achromatic designs on location charts which was recommended by Piotrowski (1957). Klopfer et al. (1954), however, did advocate questions about the position of the percept and how it was seen when probing for movement.

Among the most conservative of the Rorschach testers with regard to the inquiry are Rapaport, Gill, and Schafer (1946) who recommended conducting the inquiry after each card and with the blot not present. Questioning is kept to an absolute minimum. Thus, only what is spontaneously perceived is scored, and the record is kept free of rationalization and secondary elaboration. Sarason (1954) likewise favored a minimum of questioning, and apart from a reading back of the response with the opening query, "What was it in the card that reminded you of that?" only the clarification of any ambiguity in the client's verbalizations was allowed.

Data on the differential effects of such procedures is sparse indeed. Gibby and Stotsky (1953) have provided some information on the effects of the inquiry on Rorschach scores by comparing the scoring of the free association on patients' records to the scoring of the free association plus the inquiry. The type of inquiry used was not standardized. It was found that the indices could be differentially ordered for amount of change with *F* ranking first and *FV*, *FC*, *CF*, *M*, *YF*, *V*, and *Y* following in that order. Of course, *F* decreased while all the other indices increased, and the changes for all but *V* and *Y* were statistically significant.

Baughman (1958) proposed a new method of conducting the inquiry whereby the standard inkblots are compared systematically to a set of especially prepared blots in which color, shading, figure-ground contrast, and complexity of form are varied. *S*s' responses to a standard series of questions reveal which of the determinants have been most influential. Contrasting the scoring of records elicited by this procedure with records in which the traditional inquiry was done, Baughman

[1] The authors are grateful to Dorothea Brayer, Florence Budin, A. Cooper, J. H. Weiss, and D. L. Wolitzky who served as examiners in this study.

found that significantly more shading determinants were given in response to the paired comparison technique. However, there were no significant differences in the number of scorable color responses elicited by either technique.

In a study closely related to the one reported herein, Klingensmith (1956) investigated the effects of the amount of structure used in the inquiry on the quantity and type of indices reported. The scoring of the free association alone represented the least structured procedure. Next the scoring of the free association plus the traditional inquiry was computed. Thirdly, *Ss* were asked to return to the cards and give whatever determinants besides shape were influential in their responses and a separate scoring was done for this procedure. Finally, the determinants were described to *S* and he was told to report those he used. Again the records were scored using information elicited by this technique. When the scoring from the various procedures was compared, it was found that *T*, achromatic color, and *V* showed the greatest increase with the increase in structure while *M*, *FM*, *C*, and *Fm* showed less tendency to increase. The increase in *M* was insignificant at all levels. Unfortunately Klingensmith limited his *Ss* to two responses per card and only analyzed for 15 responses from each of the records taken.

This study is essentially a replication of that of Klingensmith with a slight procedural modification and with the entire record of each *S* being considered. This is looked upon as an exploratory study of the extent and type of change in scoring which results from a frank description of determinants concerning which the inquiry is being done. It should thereby provide data on the reliability of the indices as a function of the procedure by which they were elicited.

Method

Ss for this experiment were secured by five examiners who each administered 15 tests as part of the requirement for credit in a graduate course in Rorschach administration. Before any examiner was permitted to begin actual administration of the test, he must have participated in all of the classroom instruction and have been observed and supervised in at least two practice examinations. Examiners were instructed to use only *Ss* who were between 16 and 50 years of age and who had never been hospitalized for an emotional disorder. In the sample which was collected the age ranged from 16 to 47 years with the mean age being 25.79 years. The mean education level of these *Ss* was 14.07 years (about two years of college) and the majority (55%) were students when tested. The sample comprised 30 males and 45 females.

Procedurally, the test administration was introduced by the following statement:

> I have a set of ten inkblots that I'd like you to look at. They are made simply by dropping ink on a sheet of paper and folding the paper so that the ink spreads out in a chance way. Different people can see different things in these blots and I'm interested in what you can see in them. So, tell me whatever you see as soon as you see it. (After the first card has been returned add: From now on when you can't see any more put the card down as a signal that you've finished.)

The free association period was followed by the inquiry which was conducted in the manner described by Sarason (1954), since he gives the most precise directions and also because his approach represents one of the least structured in current use. Subsequent to this inquiry, a second inquiry was done which was prefaced by this statement:

> *Inquiry II.* The first time we went through the blots you may not have been certain just what I was getting at. The point is there are a number of features of the inkblots which may possibly help to determine the many things one can see in them. The shape or contour of the blot is obviously one of these, but other possibilities are the color of the blot, the shading within it, the suggestion of a three dimensional effect or perspective, the appearance of a textural quality to the surface and the illusion that the figures seen are in motion or assuming some position. To ensure that none of these possibilities have been missed in going over your responses, I'd like to take one more look at them and ask you to make certain that I understand in each case exactly what it was that caused the blot to look as it did to you when you first saw it. Let's go over the cards once more and you tell me what about the blot led you to see it the way you did, when you first looked at it.

In this second inquiry *Ss*' original responses and their remarks from the first inquiry were read back to them and they were invited, but not encouraged, to give any additional determinants which applied.

For each record a separate scoring was done for the free association, the free association plus the first inquiry, and the free association plus both inquiries. Approximately half of the records were scored by each of the authors, both of whom have had extensive experience with the test. Fifteen records, three being taken from each of the examiners, were scored in common to provide a reliability

check. Scoring of determinants was done according to the method prescribed by Klopfer et al. (1954) except that Beck's indices for shading were adopted.

Results

To insure the reliability of the scoring system between the two examiners a series of 27 rhos was computed. These compared the scoring of the authors on nine indices or combinations of indices elicited, for each of the three aspects of the testing situation (free association and the two inquiries). The indices studied were *M*, *FM*, *Fm*, *FC*, *CF* + *C*, *FY*, *YF* + *Y*, *FV*, and *FT* + *TF* + *T*. Seventeen rhos were greater than .90, eight fell between .80 and .90, and only two fell below .80, one being .74 and the lowest being .68 (for the scoring of *FC* on the free association in which case there were relatively few scorings made and small differences had a great effect on the overall rho). On the basis of these figures it was felt that the scoring system was a reliable one.

Since the experiment was concerned with the effect of inquiry structure on the elicitation of determinants, each *S* was used as his own control and the scorings of his free association, free association plus inquiry, and free association plus both inquiries were compared. To accomplish this, the total number of each index given by *S* in response to the entire test procedure was computed. The percentage of this total which was scorable at each of the three levels of structure was then calculated, and statistical comparisons were made between these figures. In this case the null hypothesis was that the levels of inquiry structure (the free association period was considered to represent the complete absence of such structure) had no effect on the tendency to produce the various determinants. By chance it would be expected that *S*s would either give about as many indices of any type on one phase of the test as on others or that those who give the most at a single level of structure are counterbalanced by others who are most responsive at a different level. The statistical test used was the Friedman nonparametric analysis of variance for repeated measurements of the same *S* as described by Siegel (1956).

A series of eight such analyses was done for the indices *M*, *FM*, *Fm*, *FY*, *YF* + *Y*, *FV*, *FC*, and *CF* + *C*. No analysis was done for *FT* + *TF* + *T* since so few texture responses were given. In the case of each determinant, of course, analyses were restricted to those *S*s giving them so that the *N* for each index varies somewhat. Table 1 presents the mean percentages of the indices given at each level of structure, the mean number of responses involving these indices at each level, and the χr^2 for each analysis which is referable to the standard chi square table (in this case at two degrees of freedom).

The analyses in Table 1 indicate that for all of the indices save *FV* there are highly significant differences between the percentages elicited at the various levels of structure. The findings with regard to the movement

TABLE 1

Mean Percentages and Number of Responses at Each Level of Structure and Results of Nonparametric Analyses of Variance

Index	*N*	Mean % Free Association	Mean No. of Responses	Mean % Inquiry I	Mean No. of Responses	Mean % Inquiry II	Mean No. of Responses	χr^{2*}
M	72	83	3.35	16	0.65	1	0.05	83.46
FM	73	63	3.65	32	1.88	5	0.28	62.82
Fm	41	50	1.03	48	0.99	2	0.05	24.74
FY	75	4	0.31	58	3.81	38	2.46	63.55
YF+Y	41	13	0.27	60	1.30	27	0.57	15.01
FV	33	29	0.62	48	1.05	23	0.51	2.80
FC	67	12	0.38	71	2.28	17	0.56	51.77
CF+C	60	28	0.91	62	2.05	10	0.32	39.11

* For two degrees of freedom χr^2 of 9.21 is significant at .01.

TABLE 2

TOTAL NUMBER OF *Ss* RESPONDING WITH THE VARIOUS DETERMINANTS AT EACH LEVEL OF STRUCTURE

Index	*N*	Free Association	Inquiry I	Inquiry II
M	72	68	27	3
FM	73	67	48	10
Fm	41	28	28	2
FV	75	15	67	57
VF+V	41	9	31	17
FV	33	13	20	11
FC	67	21	60	23
CF+C	60	33	51	11

categories and the shading determinants, particularly *FV*, are most striking. In the case of *M* only three *Ss* of the 72 who gave that determinant did so in response to the most structured level of inquiry, and in each case only a single *M* was added. On the other hand 60 of the 75 *Ss* ultimately giving *FV* failed to give any responses which could be scored as such on the free association, while 57 of the 75 *Ss* added this determinant in response to the instructions indicating which indices could be given. A complete breakdown of the number of *Ss* giving each index at each level of structure is found in Table 2.

DISCUSSION

The results of this study indicate that the various determinants (with the exception of *FV*) are characteristically elicited at certain times in the Rorschach examination and that these times vary with the determinant. Furtermore, they indicate that certain indices are more sensitive than others to the increase in inquiry structure. *Ss* give an average of 2.46 additional *FV* responses in the most structured inquiry while giving only 0.05 additional *M* responses in the same phase of the test. Rank ordering the indices in terms of this tendency to be elicited in the more structured inquiry, *FV* is first followed by *VF + Y*, *FV*, *FC*, *CF + C*, *FM*, *Fm*, and *M*. This is in substantial agreement with the findings of Gibby and Stotsky (1953) and Baughman (1959), who concerned himself only with color and shading. This is also in close accord with the results of Klingensmith (1956) with the exception that *FV* was not found to show the greatest increase with the increase in structure. On the contrary, of all the shade-determined indices *FV* showed the least increase on the second inquiry.

Such findings have clear implications for the clinical use of the test. They would suggest that in the absence of a standardized inquiry the *FV* index is highly unreliable. Interpretations based on the number of such determinants present in a record must be made with considerable caution therefore. On the other hand the various movement determinants seem highly reliable. In spite of the fact that what appear to be high percentages of *VF + V*, *FV*, *FC*, and *CF + C* are given on the most highly structured inquiry (see Table 1), the increment in terms of average number of such responses per *S* is never more than 0.57 which is quite small. Such differences are perhaps less than one would have been led to expect in the absence of empirical data.

While the findings in this study suggest that most scoring indices are reasonably reliable for clinical use, the implication that a standardized inquiry procedure is unnecessary is not intended. The instability of *FV* alone under varying conditions would argue for a more universally accepted way of conducting the inquiry. Perhaps one of the obstacles to the development of such a procedure is the fact that textbook writers have failed to operationalize the procedure enough. The fear that gross distortions might result from a frank confrontation of *S* with the possibilities among the various determinants may well have been a deterrent to the development of a standard set of instructions. As a result a variety of subtleties are recommended which allow for considerable personal variation in approach.

Baughman (1958) has recognized the need for greater standardization of the inquiry and has proposed a technique for eliciting color and shading responses. Unfortunately, his approach requires the use of six alternate sets of inkblots and a procedure which appears quite time consuming. The use of the kind of structuring employed in the second inquiry in this study might well accomplish the same purpose as his paired-comparison technique, and a comparison of the effects of the two

techniques would be one fruitful direction for further research.

Another worthwhile investigation would involve a comparison of the effects of examiner differences under conditions where the inquiry is an unstructured one and when, by virtue of the instructions given, it is highly structured. Conceivably, the *S* who is given few cues regarding the object of the inquiry might seize on whatever ones are inadvertently offered by the examiner, while the *S* who is given all of the cues he needs through the instructions may remain less sensitive to examiner idiosyncracies. This possibility can be investigated using the data obtained in the reported study, and the authors are currently doing so. If this hypothesis is substantiated, it would lend further support for the establishment of a standard inquiry based on a frank description of the various possible determinants to be found in the blots.

Summary

The effects of inquiry structure on the determinants elicited were studied by administering the standard Rorschach to 75 *S*s, conducting a relatively unstructured inquiry and then a second one prefaced by a description of the determinants. The free association, free association plus the first inquiry, and free association plus both inquiries were scored separately and these scorings compared. It was found that the various determinants characteristically tended to be elicited in different phases of the test procedure (i.e., most movement responses were scorable on the free association, while most shading and color responses were only scorable after an inquiry). It was also found that of the determinants studied only the change in number of *FY* on the highly structured inquiry appeared to take on clinical significance. The desirability of a standardized inquiry was discussed, and the possibility of basing it on a frank description of the determinants which are the object of inquiry was considered.

REFERENCES

Baughman, E. E. A new method of Korschach inquiry. *J. proj. Tech.*, 1958, **22**, 381–389.

Baughman, E. E. The effect of inquiry method on Rorschach color and shading scores. *J. proj. Tech.*, 1959, **23**, 3–7.

Beck, S. J. *Rorschach's test.* Vol. 1. *Basic processes.* New York: Grune & Stratton, 1950.

Gibby, R. G., & Stotsky, B. A. The relation of Rorschach free association to inquiry. *J. consult. Psychol.*, 1953, **17**, 359–364.

Klingensmith, S. W. A study of the effects of different methods of structuring the Rorschach inquiry on determinant scores. Unpublished doctoral dissertation, Univer. Pittsburgh, 1956.

Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach technique.* Vol. 1. *Theory and technique.* New York: World Book, 1954.

Klopfer, B., & Kelley, D. M. *The Rorschach technique.* New York: World Book, 1946.

Piotrowski, Z. A. *Perceptanalysis.* New York: Macmillan, 1957.

Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Vol. 2. Chicago: Year Book Publishers, 1946.

Sarason, S. B. *The clinical interaction.* New York: Harper, 1954.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received August 19, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 333-341

CLINICAL DIAGNOSIS BY THE IPAT MUSIC PREFERENCE TEST

RAYMOND B. CATTELL AND ROBERT E. McMICHAEL
University of Illinois

The present investigation was undertaken with the limited purpose of determining any existing differences in the types of musical preference expressed by a normal control group and various pathologic, psychotic, and neurotic groups. It is an exploratory piece of research in that the measuring instrument used is of quite recent vintage. Though it is convenient, consisting of a single long-playing record which can be responded to on answer sheets by a whole collection of patients, and although the independence of the factor scales has been factor analytically confirmed, very little information has yet come forth regarding its validity or reliability as a clinical instrument.

A search of the literature has shown that even though several studies have employed measures of musical preference not one has directly concerned itself with investigating the determination of diagnostic classification of psychiatric syndrome groups by musical preference. Rather, they have been principally concerned with determining possible variables influencing or associated with musical preference in *normal* subjects (*Ss*). Among others, such variables as age (Keston & Pinto, 1955; Valentine, 1926), sex (Fay & Middleton, 1941; Kerr, 1943; Keston & Pinto, 1955; Schultz, 1933; Sopchak, 1955), mood (Coroso, 1958; Podolsky, 1954; Sopchak, 1955), "intellectual introversion" (Keston & Pinto, 1955), and music recognition (education) (Keston & Pinto, 1955; Sopchak, 1955), have been investigated. The results of such studies at best suggest some variables that may profitably be controlled when working with measures of musical preference.

In a more dynamic mode, both Cattell (1951, 1957) and Sopchak (1957) have proposed the use of music as a "projective" device. The former starts with the premise that, "The powerful and immediate connection of musical stimulation with emotional experience, and the many indications that unconscious needs gain satisfaction through this medium, have long pointed to measures of musical preference as effective avenues to deeper aspects of personality" (Cattell & Anderson, 1953b, p. 446). Believing that here, as in his humor studies, psychiatric judgment is not sufficient to justify a priori decisions as to what musical preferences should form the scale content for a particular syndrome, he factored the likes and dislikes for 120 brief musical excerpts and arrived at 11 independent (simple structure) scale dimensions (Cattell & Saunders, 1954). These cannot be explained in terms of historical musical periods or schools but suggest dynamic factors of temperament and emotional need. On this research basis a 100-item, 1-hr. test was recorded and standardized as described elsewhere (Cattell & Anderson, 1953a; Cattell & Anderson, 1953b). Subsequently, a correlation study was carried out between the music factors of the musical preference test and the factors on the 16 Personality Factor Questionnaire (Cattell & Anderson, 1953b). The study showed that although a simple one-to-one relation between the two sets of factors could not yet be established, the correlations are substantial, i.e., the test gets into the same "personality space," and, "the set of 16 PF test factors associated with any one music factor has a psychologically consistent and compatible character . . . in every case" (Cattell & Anderson, 1953b, p. 449).

The same study reports results of music factor score comparisons among psychotics and normals, and it is this part with which the present study continues to further cat-

egories. Such comparisons were suggested by reasoning that if the musical preferences are functions of personality factors then various psychotic and neurotic syndrome groups should show different music factor scores. This reasoning is based on the commonly accepted—however, often untested—proposition that various syndromes are interpretable in terms of extreme patterns of common personality factors (as shown, for example, in Cattell and Scheier's recent studies [1960] of neurotic syndromes with the 16 PF test). Statistically significant results were in fact found, and on the basis of this suggestive, promising, and partially confirmatory evidence of the test's capacity to discriminate among pathologic groups, it was concluded that further investigation, using other samples, was desirable. Consequently, in overall design the present research is similar to that of Cattell and Anderson (1953b). However, because of modifications in the diagnostic categories and sizes of the pathologic groups, in testing procedure and statistical analysis, and in the sex, age, and other characteristics of all the samples, it is by no means a straightforward replication.

Method

Design

The IPAT Musical Preference Test, permitting scores on 11 distinct dimensions of choice, was administered to several diagnostically different abnormal groups, in addition to a control or "normal" group. Intergroup comparisons, on their mean scores on each factor, were made between (*a*) the various pathological groups and the control group, (*b*) the control group and the *total* abnormal group, and (*c*) among the various pathological groups. The design included the additional features that only *same sex* comparisons were made and that any possible effects attributable to age were first separated out. These comparisons provided the basis for conclusions concerning the discriminating power of the music preference test.

Subjects

The *Ss* of this investigation consisted of seven groups—organics, sociopathic personalities, schizophrenics, paranoids, affectives, psychoneurotics, and normals. Table 1 presents general, descriptive, identifying data on all the samples used in this investigation. All of the abnormal *Ss* were patients in Manteno State Hospital, Manteno, Illinois. The criterion used for diagnostic classification was the hospital diagnosis as determined in case conferences. A restriction imposed on the abnormal samples was that no severely regressed patients participated in the study. Every patient was able, at least initially, to attend to the music.

Since it seemed undesirable to take students as a control group, as was done in an earlier study, we obtained the cooperation of members of a church social group of roughly similar social status and age range to the mental hospital group. Descriptive data on this sample are, however, limited because these *Ss* were not required to list any identifying data other than their age and sex. However, educationally, they seemed somewhat more homogeneous, though not necessarily higher, than the abnormals.

Materials

The IPAT Music Preference Test of Personality (Cattell & Anderson, 1953a) was the test instrument used in the present investigation. The test consists of 100 musical excerpts which have been put onto a 12-in., 33⅓ rpm record by RCA Victor. The test was tape recorded in order to facilitate group testing. Good fidelity of reproduction was obtained. A Knight tape recorder and an external speaker system were used for the playing of the music.

Testing Procedure

The normal *Ss* were tested in one group in a large room at the church. The abnormal *Ss* were tested in groups as formed by wards. Therefore, various diagnostic groups were represented at a given testing. The taking of the test was strictly a matter of voluntary participation on the part of the patients. There were always from 8 to 10 monitors present who encouraged and supervised the *Ss*, showing them where to record their choices if they appeared to need help. Talking during the test was generally not permitted. The total testing time varied but usually was about one hour.

Scoring

Scoring of the answer sheets was done by using the answer keys provided in the test materials. These keys yield total "weighted item" raw scores on each of the 11 musical factors which together comprise the music preference test. Once these weighted raw scores were obtained from every *S*'s answer sheet, the actual scoring was completed. These score values are the data used in the statistical treatments.

Although a total of approximately 300 patients were present on the wards, only 230 actually participated and submitted complete answer sheets. Of these 230 complete records, 190 were usable. Forty records had to be discarded for one reason or another, e.g., filling all the blanks out before the test actually began, copying the responses of one's neighbor, attending to the music intermittently, and so forth. Since the diagnostic labels were obtained from the hospital only for those patients who met the testing criterion and had usable answer sheets, the diagnoses of these 40 patients are not known.

TABLE 1

DESCRIPTIVE IDENTIFYING DATA ON SAMPLES

Sample Type	Sample Size	Age in Years		No. of Months in Hospital		Educational Level in Grades		No. of Re-admissions		% in Some Type Treatment	% of Voluntary Commitment	%of Non-white	No. *Ss* of Different Religions[a]			No. *Ss* of Given Marital Status[b]			
		Mean	Range	Mean	Range	Mean	Range	Mean	Range				P	C	O	M	S	D	O
Males																			
Organics	30	46.4	24–67	20.7	0–108	8.6	0–16	1.8	0–6	57	37	17	14	15	1	8	9	7	6
Schizophrenics	19	34.4	21–53	47.6	1–180	8.8	0–16	.8	0–5	74	0	42	10	77	2	6	10	2	1
Sociopathic Personalities	25	47.3	22–74	11.1	0–110	9.5	0–16	2.1	0–12	60	76	8	11	14	0	3	3	11	8
Paranoids	11	51.6	42–63	75.3	1–242	9.1	8–13	.7	0–4	73	0	18	3	7	1	7	1	2	1
Affectives	10	48.9	18–58	32.5	5–146	6.8	0–12	1.3	0–4	30	30	0	2	6	2	6	2	2	0
Psychoneurotics	9	31.8	15–63	92.3	2–255	8.3	6–11	.1	0–1	33	11	0	3	2	4	0	8	0	1
Total Abnormals	104	44.0	15–74	46.6	0–255	8.5	0–16	1.1	0–12	58	33	16	43	51	10	30	33	24	17
Normals	26	57.5	49–66									0	26	0	0				
Females																			
Organics	13	41.5	25–62	9.7	1–111	8.4	0–12	2.2	0–9	77	15	8	8	5	0	4	3	2	4
Schizophrenics	32	39.6	21–63	53.7	3–139	10.1	0–14	.8	0–6	62	3	31	14	14	4	10	9	4	9
Sociopathic Personalities	8	45.2	26–64	17.2	1–75	8.1	0–13	1.1	0–5	50	25	25	4	4	0	2	1	1	4
Paranoids	10	55.7	35–74	71.1	5–252			.4	0–2	40	10	20	4	5	1	4	3	2	1
Affectives	15	42.0	26–58	41.7	4–202	8.1	0–12	2.3	0–6	60	0	7	4	11	0	7	4	0	4
Psychoneurotics	8	45.5	16–62	21.9	2–75	8.8	7–16	4.1	0–9	75	50	0	5	3	0	3	2	2	1
Total Abnormals	86	43.3	16–74	35.9	1–252	8.7	0–16	1.8	0–9	62	12	19	39	42	5	30	22	11	23
Normals	34	55.2	45–66									0	34	0	0				

[a] P = Protestant; C = Catholic; O = Other (Jewish, none, or not known).
[b] M = Married; S = Single; D = Divorced; O = Other (separated, widowed, or not known).

Statistical Treatment of Results

The statistical treatment of the results involved three steps, each step designed to determine whether or not a certain variable or category had an influence on scores of music preference. First, an analysis of covariance was carried out on each of the 11 music factors involved in the test. This analysis was employed in order to demonstrate that a classification based on diagnosis and sex would show significant effects when effects attributable to age were removed. Second, on those factors for which the classification showed significant effects, a two-way analysis of variance (sex by categories) was carried out to ascertain whether significant score differences should be attributed to sex, diagnosis, and/or interaction. Finally, contingent upon the results of the analyses of variance, *t* tests were carried out to determine the *location* of statistically significant differences among the diagnostic groups. It was also planned that the analyses of variance and *t* tests should be carried out on scores adjusted for regression on age.

Results

An analysis of covariance was run on each of the 11 music factors. Consequently, any effect of age on the music preference scores was statistically controlled for. Out of the 11 analyses of covariance, 7 were significant at the 10% level of significance agreed to justify further analysis.[1] Although this was the value of significance agreed to warrant further analysis, the probabilities associated with all but one of the *F* ratios were at the 5% level or less. For these 7 factors on which such a result was obtained it may be said that the variance among the means of the 14 groups, i.e., 7 categories by two sexes, was large enough to warrant further analysis. On the remaining 4 factors for which the null hypothesis was not rejected, no further analysis was undertaken.

An analysis of variance was run on each of the seven music factors which showed a significant analysis of covariance result. For purposes of exploratory analysis, if the *F* yielded by the analysis of variance had a chance probability of .10 or less, it was to be considered a statistically significant value. Though this was the level of statistical significance adopted by the authors, the actual *F* ratios for the main effect of diagnosis had associated probabilities of .05 or less.

The seven analyses of variance had in common the result of a significant main effect due to diagnostic category. The findings showed that for a given factor this effect might be the only statistically significant one or that it occurred in addition to a main sex effect or an interaction effect. A main sex effect was found on Factors 1, 8, and 11, indicating that a consistently higher or lower score is obtained by a given sex regardless of diagnostic group membership. Using observed score values, an interaction effect was found on Factors 2 and 6, indicating a variable higher or lower score by a given sex depending on group type. (These interactions drop out, however, if adjusted scores are used.) Because sex effects were present on these factors, only same-sex comparisons were permitted in *t* tests of mean comparisons, as stated earlier. Although sex effects did not appear in the analysis of variance of Factors 5 and 10, the next part of the analysis proceeds, for the sake of clarity and consistency, by cautiously assuming that they might.

The *t* tests of mean comparisons were then run in order to determine just where the statistically significant differences due to sample type occur. Except for Factors 1 and 11, the means used in the comparisons were the observed means of the samples, not means after adjustment for age, since inspection of the regression coefficients indicated that adjusting for age was not necessary, save on Factors 1 and 11. In addition, no multiple comparison technique seemed justified since it typically involves a within-groups or pooled-variance expression, and this was not compatible with the heterogeneity of variance which exists in some of the comparisons. Therefore, independent estimates of the population variance were used in the calculation of every *t* ratio and the degrees of freedom associated with each directly calculated (see Walker & Lev, 1953, pp. 157–158). The main consequence

[1] Due to limitations of space, the summary tables of raw score means and variances (Tables A and B), the summary tables of the analyses of covariance and variance (Tables C and D, respectively), and the summary table of regression coefficients (Table E) are not presented in this report, but they have been deposited with the American Documentation Institute. Order Document No. 6331 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 × 8 in. photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2

THE *t* VALUES OF DIFFERENCES OF CLINICAL GROUPS FROM CONTROLS, ONE ANOTHER, AND TOTAL ABNORMAL GROUP FOR MALES

Significance of Differences	Factors						
	1	2	5	6	8	10	11
Controls vs.							
Total Abnormals	1.1	4.2***	−1.1	−3.3***	5.1***	2.1**	−3.0***
Organics	0.7	4.8***	−1.8*	−1.8	3.6***	2.0**	−1.6
Sociopathic Personalities	0.3	2.7**	−1.6	−1.3	1.9	2.0	−1.0
Schizophrenics	1.4	2.4**	0.6	−2.8***	4.9***	0.5	−2.2**
Paranoids	−1.4	2.5**	−1.0	−4.3***	2.6**	2.4**	−0.3
Affectives	0.2	0.2	−0.8	−1.0	1.3	0.2	−1.0
Psychoneurotics	4.2	2.4**	0.8	−2.4**	5.9***	2.1**	−3.3***
Total Abnormals vs.							
Organics	−0.2	2.0**	−1.4	0.8	−0.4	0.1	0.8
Sociopathic Personalities	−0.7	−1.3	−1.1	2.0	−2.1**	0.6	1.2
Schizophrenics	0.9	0.5	1.9*	−0.7	1.7	−1.3	−0.7
Paranoids	−2.8	0.5	−0.6	3.0**	−0.3	1.3	0.9
Affectives	−0.5	−2.4**	−0.2	0.9	−0.7	−1.0	0.0
Psychoneurotics	4.3***	−0.4	1.5	−0.9	3.6***	1.0	2.5**
Organics vs.							
Sociopathic Personalities	−0.3	−2.3**	0.1	0.7	−1.2	0.4	0.4
Schizophrenics	0.8	−0.3	2.1**	−0.9	1.5	−1.0	−1.0
Paranoids	−2.1**	−0.3	0.1	−2.7**	0.0	1.1	0.5
Affectives	−0.3	−2.8**	0.4	0.3	−0.5	−1.0	−0.2
Psychoneurotics	3.7***	−1.4	1.8	−1.1	3.2**	0.9	−2.4**
Sociopathic Personalities vs.							
Schizophrenics	1.0	0.9	1.9*	−1.6	2.4**	−1.2	−1.2
Paranoids	−1.5	1.0	0.0	−3.4***	0.9	0.7	0.2
Affectives	0.0	−1.6	0.4	−0.1	0.2	−1.1	−0.5
Psychoneurotics	3.6***	0.4	1.7	−1.6	3.9***	2.3**	−2.6**
Schizophrenics vs.							
Paranoids	−2.6**	0.0	−1.4	−1.8	−1.2	1.7	1.1
Affectives	−0.8	−1.9	−1.2	1.0	−1.3	−0.2	0.4
Psychoneurotics	2.6**	−0.6	0.3	−0.4	2.0	1.5	−1.7
Paranoids vs.							
Affectives	1.2	−2.0	0.3	2.6**	−0.4	−1.5	−0.6
Psychoneurotics	5.1***	−0.6	1.4	1.1	2.8**	−0.1	−2.4**
Affectives vs.							
Psychoneurotics	3.0***	1.7	1.3	−1.2	2.6**	1.4	−1.7

Note.—A minus value indicates that the group preceding "vs." has the higher mean.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

of this procedure is in keeping with conservative presentation of results since the error involved is one of obtaining fewer significant *t* ratios than if a multiple comparison technique were employed.

In the following presentation of the *t* test results it seemed advisable merely to set forward the findings at a descriptive level without any attempt at deeper psychological interpretation. This approach seems most appropriate considering the exploratory level and intent of this investigation. It also seems recommended

by the fact that, although the factors are well established empirically, their psychological interpretation is necessarily tentative.

Table 2 presents the significance of differences of the clinical groups from the controls, from one another, and from the total abnormal group, for the males only. Table 3 presents the same results for the females. No t value having a probability greater than 5% was considered noteworthy except for such an occurrence on Music Factors 1 and 5. On these factors a probability of 10% was con-

TABLE 3

THE t VALUES OF DIFFERENCES OF CLINICAL GROUPS FROM CONTROLS, ONE ANOTHER, AND TOTAL ABNORMAL GROUP FOR FEMALES

Significance of Differences	Factors						
	1	2	5	6	8	10	11
Controls vs.							
Total Abnormals	0.4	5.3***	−0.4	−5.6***	5.7***	3.0***	−1.8
Organics	0.2	1.6	−2.0*	−2.1	2.9***	2.5**	0.2
Sociopathic Personalities	−2.3**	2.6**	−1.0	−0.4	1.1	1.3	0.3
Schizophrenics	0.1	3.5***	−0.5	−6.1***	4.6***	3.0***	−3.0***
Paranoids	1.2	3.3***	2.1*	−1.7	1.8	2.8**	0.5
Affectives	0.1	3.8***	0.6	−4.2***	4.0***	0.3	−2.0
Psychoneurotics	0.3	1.5	−0.3	−2.4**	2.4**	1.1	0.8
Total Abnormals vs.							
Organics	−0.1	−1.2	−2.1**	0.4	0.0	0.7	1.4
Sociopathic Personalities	−3.5***	0.0	−0.9	3.5***	−1.5	−0.1	1.2
Schizophrenics	2.0**	−0.6	−0.2	−1.9	1.3	0.5	−2.3**
Paranoids	1.1	0.7	2.7**	1.6	−1.9	1.1	1.9
Affectives	−0.4	1.3	1.0	−1.3	0.6	−2.4**	−1.3
Psychoneurotics	0.0	0.0	−0.2	0.3	−0.1	0.2	1.7
Organics vs.							
Sociopathic Personalities	−1.8*	0.7	1.0	1.9	−1.1	−0.4	0.2
Schizophrenics	0.6	0.7	2.9***	−0.9	0.5	−0.3	−2.1**
Paranoids	0.8	0.2	4.7***	0.7	−1.2	0.4	0.2
Affectives	−0.1	1.6	2.6**	−1.0	0.3	−2.0	−1.7
Psychoneurotics	0.1	0.5	1.0	−0.1	−0.1	−0.1	0.6
Sociopathic Personalities vs.							
Schizophrenics	3.0**	−0.2	0.7	−3.9***	1.8	0.2	−1.8
Paranoids	2.9**	0.4	2.5**	−1.1	0.2	−0.7	0.0
Affectives	2.1*	0.9	0.3	−3.2	1.5	−1.4	−1.5
Psychoneurotics	2.1**	0.0	0.3	−1.8	1.0	0.2	0.4
Schizophrenics vs.							
Paranoids	0.4	0.8	2.3**	2.3**	−2.0	0.7	2.7**
Affectives	−0.9	1.3	0.8	−0.2	−0.2	−2.1**	0.0
Psychoneurotics	−0.5	0.2	0.0	0.9	−0.6	0.1	2.2**
Paranoids vs.							
Affectives	−1.0	0.4	−0.1	−1.9	1.7	−2.3**	−2.1**
Psychoneurotics	−0.8	−0.3	−1.6	−0.8	1.0	−0.3	0.4
Affectives vs.							
Psychoneurotics	0.2	−0.6	−0.6	0.9	−0.4	1.0	2.0

Note.—A minus value indicates that the group preceding "vs." has the higher mean.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

sidered statistically significant. This seemed in keeping with the 10% level of significance that was accepted in the analysis of variance on Factors 1 and 5. Three levels of significance are reported merely in order that a given reader may consider those values which suit his particular research requirements, not because the authors have three levels of confidence in the present results.

Tables 2 and 3 reveal that the control group differs significantly from each of the pathologic groups on at least one factor and often on five or six. Tables 2 and 3 also present the results which indicate the ability of the seven music factors to discriminate among the various syndrome groups themselves. Examination of Table 2 reveals that the male organics show a significant difference, when compared with the total abnormal group, on Factor 2. They are also distinguished from each of the various other diagnostic groups by a difference on some one factor. The female organics are more easily distinguished from the other groups since they differ most on only one factor. Their lower score on Factor 5 distinguishes them from every group except the neurotics and sociopathic personalities.

The male sociopathic personality group shows no difference from the affectives. However, they are distinguished from all the remaining groups by a significant difference(s). The female sociopathic personalities differ from the total abnormals, schizophrenics, and affectives by being higher on Factor 6, and from the paranoids by being lower on 5. In addition, they differ significantly from the total abnormal and all the other pathologic groups by scoring lower on Factor 1.

The male schizophrenic sample is distinguished from the organics by a higher score on 5, and also from the sociopathic personalities by a higher score on Factors 5 and 8. They are also characterized by scoring higher on Factor 1 than paranoids, and lower than psychoneurotics. The female schizophrenics show a greater number of distinguishing comparisons. Possibly Factor 11 will prove to be the most discriminatory of this reaction.

The lower score of the male paranoids on Factor 6 distinguishes them from all other groups except for the neurotics and schizophrenics. The psychoneurotics score higher, while the schizophrenics score lower, on Factor 1 than the paranoids. They show no other difference from the schizophrenics, but have a lower score on Factor 8 than the neurotics. The female paranoid sample is principally characterized by being significantly higher on Factor 5 than most of the other groups.

A higher score on either or both Factors 1 and 8 distinguished the male psychoneurotics from the other samples. Moreover, a significantly lower score on Factor 11 is obtained by the neurotics except in comparison to the schizophrenics and affectives. The female psychoneurotics have no consistent significant differences from the other clinical groups on any given factor.

The male affective group differs significantly from organics, paranoids, and psychoneurotics on some factors, as have already been indicated. The female affective group differs from the organics by being higher on Factor 5, from the sociopathic personalities by being lower on Factor 6, higher on 1, and from both schizophrenics and paranoids by being lower on 10. They also differ from paranoids by being lower on 11.

To summarize these differences in terms of those which hold for both sexes—and these are of most value for the practicing clinician —we find:

1. Abnormals are higher on Factors 2, 8, and 10, and lower on 6 (and 11, if one goes to the 5% level). Among these, organics are higher only on Factors 2 and 8; neurotics on 8; and schizophrenics on 2, 6, 8, and 11.

2. Sociopaths differ from all other abnormals in being high on Factor 6.

3. Schizophrenics differ, notably from organics, in being high on Factor 5 and relatively low on 11.

4. Neurotics distinguish from most groups by being higher on Factor 1 and, less definitely, on Factor 8.

5. Paranoids are higher on Factor 11, and affectives are lower on 11, than most.

Since the component factor scores on these patterns are virtually independent, it will be seen that multiple correlations between the specific diagnostic category and the scores on the component factors in the different patterns could well be appreciable, but such finer cal-

culation must be left until confirmation is available on other samples. Meanwhile it would seem that 7 of the 11 factors in this test are important for clinical diagnosis.

Discussion

Although, as stated, the Music Preference Test factors are at a descriptive rather than an interpretive stage, some discussion of their associations with abnormality may be profitable.

The descriptions are at present based on 16 PF test correlations (Cattell & Anderson, 1953b) and for the seven factors involved here are as follows:

F1. Adjustment (vs. Frustrated Emotionality) which correlates negatively with ergic tension (Q_4).
F2. Parasympathetic Immunity vs. Sensitivity to Threat.
F5. Introspectivness vs. Social Contact, having a substantial correlation with I factor (see Cattell, 1957).
F6. Anxiety vs. Paranoid Imperviousness, correlating with O factor (guilt proneness).
F8. Resilience vs. Withdrawn Schizothymia (H factor).
F10. Lack of Drive.
F11. Tenacity vs. Cyclothyme Relaxedness.

Some "face validity" is given to these descriptions from the *normal* range, when we move to the present *abnormal* sample, in several respects. First, paranoids are higher on "paranoid imperviousness," and schizophrenics differ from other abnormals in being high on Factor 5, introspectiveness. However, abnormals generally are marked by toughness, i.e., parasympathetic predominance and resilience, plus lack of drive and lack of guilt proneness. The neurotics score more "adjusted," Factor 1, than the psychotics, and more resilient. Affectives show more "cyclothyme relaxedness," Factor 11(–).

The last relation was also found in the Cattell & Anderson study (1953b), as well as the higher scores of schizophrenics on introspectiveness; but, except for schizophrenics and affectives the studies are not comparable in their groups. Although this much consistency exists, there are several other instances—such as the high score of sociopaths on anxiety, Factor 6, relative to other abnormal groups—which are relatively inapt and suggest that we are only just beginning to grope our way to the true meaning of these music preference factors. Some similar oddities of label have been noted by Coroso (1958). This is not surprising considering how little has yet been done to relate the factors to other landmarks in personality measurement. Despite our inability to interpret, however, the fact remains that the test has on both occasions of experiment shown highly significant differences of abnormals from controls and among abnormal, psychiatric syndrome groups.

Age as a possible influencing variable in the present study with abnormals was shown to have no appreciable relation to music preference. This seems to be in agreement with studies which reported the same finding, although they used a different type of measurement and *normal Ss* (Keston & Pinto, 1955; Valentine, 1926). Theoretically, one might argue that such results are confirmatory evidence for music preferences being considered more a function of the temperament of the perceiver than a function of experience, transient needs, familiarity, and musical training, which vary with age. Perhaps temperament largely determines the music one likes and becomes familiar with.

Questions concerning the influence of such variables as race and marital status on the music preference factor scores are difficult to answer from the present samples. Insofar as an answer is possible concerning racial differences, analysis with nonwhites eliminated indicates that it makes no difference. However, these are variables requiring further investigation.

Summary

An exploratory study in the general area of music preference and personality diagnosis was undertaken for the purpose of searching out music preference differences of various psychiatric samples.

The IPAT Music Preference Test of Personality was administered to a group of controls and to six "diagnostically" different groups of abnormal patients. The test was scored in standard fashion to yield scores on each of the 11 music factors comprising the

test. Previous researches in the area had shown age and/or sex might influence music preference. Consequently their possible influence was examined in the present study.

The results showed that age is not significantly related to music preference but that there are sex differences associated with music preference. Subsequently all music preference comparisons were made within the same sex. Seven out of the original 11 test factors yielded statistically significant t tests of comparisons. Several of the music factor scores obtained by various groups were considered in terms of their psychological meaning, as deduced from personality test correlations in normals, with the result that some designations seem appropriate and others in need of revision.

It was concluded that the Music Preference Test of IPAT holds promise as a diagnostic instrument, as indicated by the marked score differences between various groups. The conclusion was also made that the test is best employed by treating results for males and females separately. Because of limitations of the present investigation it is believed that further studies should be carried out in order to determine whether or not the relations identified in this research are replicable.

REFERENCES

CATTELL, R. B. Principles of design in "projective" or misperception tests of personality. In H. H. Anderson & G. H. Anderson (Eds.), *Projective techniques*. New York: Prentice-Hall, 1951.

CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.

CATTELL, R. B., & ANDERSON, J. C. *The I.P.A.T. Music Preference Test of Personality*. Champaign, Ill.: IPAT, 1953. (a)

CATTELL, R. B., & ANDERSON, J. C. The measurement of personality and behavior disorders by the IPAT Music Preference Test. *J. appl. Psychol.*, 1953, **37**, 446–454. (b)

CATTELL, R. B., & SAUNDERS, D. R. Musical preferences and personality diagnosis. I. A factorization of 120 themes. *J. soc. Psychol.*, 1954, **39**, 3–24.

CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press, 1960.

COROSO, JOAN. Relationship between the IPAT Music Preference Test of Personality and E, S, and T scores on the Guilford-Zimmerman temperament survey. Unpublished master's thesis, Univer. Connecticut, 1958.

FAY, P. J., & MIDDLETON, W. C. Relationship between music talent and preference for different types of music. *J. educ. Psychol.*, 1941, **32**, 573–585.

KERR, W. A. Three studies in plant music. *Industr. Music News*, 1943, **1**, 4–10.

KESTON, M. J., & PINTO, I. M. Possible factors influencing musical preference. *J. genet. Psychol.*, 1955, **86**, 101–113.

PODOLSKY, E. (Ed.) *Music therapy*. New York: Philosophical Library, 1954.

SCHULTZ, F. J. Testing listening power in music. *Music Supervisors' Nat. Conf. Yearb.*, 1933, **26**, 306–312.

SOPCHAK, A. L. Individual differences in reactions to different types of music, in relation to sex, mood and other variables. *Psychol. Monogr.*, 1955, **69** (11, Whole No. 396).

SOPCHAK, A. L. Relation between MMPI scores and musical projective test scores. *J. clin. Psychol.*, 1957, **13**, 165–168.

VALENTINE, C. W. *An introduction to the experimental psychology of beauty*. New York: Dodge, 1926.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received August 24, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 342–348

PERSONALITY CORRELATES OF MANIFEST ANXIETY IN CHILDREN[1]

LUCIANO L'ABATE

Washington University School of Medicine, St. Louis

Wirt and Broen (1956) have questioned whether any relationship exists between anxiety as measured by the Children's Manifest Anxiety Scale (CMAS) (Castaneda, McCandless, & Palermo, 1956; McCandless & Castaneda, 1956) and the concept of anxiety as used in the clinic. Such a lack of relationship may be confounded by sex differences in defending against anxiety. Although boys and girls may obtain similar scores on tests like the CMAS, they may use different defenses in coping with anxiety. Such a viewpoint assumes that anxiety is a variable, with deviations in behavior viewed as defenses against it. If the assumption is correct, as reviewers of emotional sex differences in children lead one to believe (Kuhlen, 1957, p. 274; Terman & Tyler, 1954, p. 1075), one would predict that different relationships would hold between the CMAS and various measures of adjustment in boys and girls.

Psychoanalytic writers (Bonaparte, 1953; Freud, 1946; Greenacre, 1948; Silverberg, 1952) have broadly suggested that defenses against anxiety may be divided into two major categories that are pertinent to the assumption underlying the present study, namely: girls handle anxiety through autoplastic defenses, while boys handle it through alloplastic defenses. One would predict that although boys and girls may report the same amount of anxiety on a paper and pencil test like the CMAS, the relationships between their scores on such a test and various measures of adjustment may vary considerably. Measures that could be divided rationally into autoplastic and alloplastic defenses should also differentiate between boys and girls. Although maladjustment as a total measure may represent the combination of autoplastic and alloplastic defenses, girls should show a greater degree of relationship between manifest anxiety scores and autoplastic defenses such as daydreaming and feelings of inferiority, while boys should show a greater degree of relationship between manifest anxiety scores and alloplastic defenses, such as rebelliousness and general "nuisance value." No more specific predictions than the above could be made because of difficulties inherent in classifying measures of adjustment as being clearly either autoplastic or alloplastic.

The aims of the present investigation, therefore, are twofold, namely: to study whether sex differences exist in the management of anxiety as well as to specify the nature of such differences by correlating the CMAS with Rogers' Test of Personality Adjustment (Rogers, 1931a) and Rogers' Rating Scale of Adjustment (Rogers, 1931b).

PROCEDURE

Sample

Forty-nine boys and 47 girls in the fourth to eighth grades (9 to 13 years old) were administered the CMAS and the Rogers' Test of Personality Adjustment (1931a) in groups of 20 subjects (*Ss*) or less. There were 10 boys and 10 girls in each grade except for the fourth grade, which contained 9 boys and 7 girls. The *Ss* were selected from a public school population on the basis of their average or above-average IQ. The mean Otis Self-Scoring IQ for boys was 106.38, $SD = 8.22$, while the mean Otis Self-

[1] An earlier version of this paper was read at APA Convention, Cincinnati, Sept., 1959. The author acknowledges the cooperation of Pitt County Health Dept., and of the Board of Education, Greenville, N. C. The data of this investigation were collected by Laura Ciporin through the cooperation of Edna Baker, who also collected the teachers' ratings. This report was written during the author's tenure as USPHS postdoctoral fellow at the Institute for Psychosomatic and Psychiatric Research and Training of Michael Reese Hospital, Chicago, Illinois.

Scoring IQ for girls was 112.21, $SD = 9.33$. The background of the sample was strictly rural, lower to middle class economically. Their school was located in a town of 3,000 in eastern North Carolina.

To determine whether our North Carolina sample was similar to the original Iowa City sample in terms of manifest anxiety level and the number of lies (the score for lies is derived from 11 questions that buffer the CMAS), the means of our fourth, fifth, and sixth grade boys and girls were compared to Levitt's (1957) data. He compared the original rural sample to an urban (Chicago) sample, finding a much lower mean level of manifest anxiety in urban than in rural children (Table 1). Except for a significantly greater Lie score in boys, our North Carolina sample did not differ significantly from the original Iowa City sample in terms of anxiety level.

TABLE 1

ECOLOGICAL DIFFERENCES ON THE CHILDREN'S MANIFEST ANXIETY SCALE BETWEEN NORTH CAROLINA, IOWA CITY, AND CHICAGO SAMPLES FOR FOURTH, FIFTH, AND SIXTH GRADE BOYS AND GIRLS

	Anxiety					
	Boys			Girls		
	Mean	*SD*	*N*	Mean	*SD*	*N*
North Carolina	17.57	6.13	29	18.09	6.25	27
Iowa City	15.95	7.57	202	18.42	8.36	184
Chicago	13.59	7.56	161	12.87	8.23	161
	Lies					
North Carolina	3.23	2.08	29	3.29	2.41	27
Iowa City	2.02	1.85	187	2.52	2.11	174
Chicago	3.92	2.42	161	4.29	2.30	161

Significance of Differences between Means

	Anxiety			
	Boys		Girls	
	t	*p*	*t*	*p*
North Carolina vs. Iowa City	1.10	ns	.02	ns
North Carolina vs. Chicago	2.69	.01	3.14	.01
	Lies			
North Carolina vs. Iowa City	3.54	.001	1.71	ns
North Carolina vs. Chicago	1.31	ns	2.08	.05

Furthermore, our sample differed from the Chicago sample to the same extent as the Iowa City sample did (Table 1).

In view of the above, conclusions and generalizations based on our sample should apply only to comparable samples of rural children. Generalizations about the significance of our findings to urban children should be made with caution, because of the significant differences in anxiety level existing between urban and rural children. Although the ecological comparison was made with children between the fourth and sixth grades, it should be noted that our sample also included children from seventh and eighth grades, grades not included in the original standardization of the CMAS. The two higher grades were included because of the applicability of Rogers' Test to a wider age range than the CMAS.

Description of Criterion Measures

The Rogers' Test of Personality Adjustment is a paper and pencil questionnaire filled out by the child himself. It yields separate scores for personal inferiority, social maladjustment, familial maladjustment, and daydreaming. The four scores and their sum are so directed that the higher the numerical score, the greater the degree of maladjustment. Besides Rogers' original validation (1931a, 1931b), a recent study by Smith (1957) showed that the test differentiated satisfactorily among three groups of well adjusted, adjusted, and maladjusted boys in sixth grade. The means of our 49 boys on the Rogers' Test compare well by inspection with Smith's sample (Smith, 1957, Table 22). The sole exception may be the daydreaming score, which appears to be slightly higher in Smith's all urban sample.

The scores achieved on the Rogers' Test by our public school sample were compared to those of 38 children referred to a mental health clinic for behavior problems. The clinic children were of the same age span, intelligence, and socioeconomic and geographic background as the normal school children. There were more clinic boys (34) than clinic girls (4), the latter being an admittedly small sample. The mean scores of clinic children on Rogers' Test were significantly greater on all four subscales than the public school children, except for family maladjustment scores for boys and social maladjustment scores for girls (Table 2). There were no statistically significant differences between the means of clinic boys and clinic girls. Either in differentiating a pathological sample of children from normal children, or in comparing our sample with a normal sample of sixth grade boys, Rogers' Test of Personality Adjustment appears to possess sufficient validity to be used as an internal criterion measure for the CMAS.

As an external criterion measure, each child was rated by two different teachers on Rogers' Rating Scale of Adjustment (1931b). All ratings are so directed that a high score indicates a greater degree of maladjustment. For instance, a score of 1 on the scale of independence means "very independent, anxious to grow up," while a score of 5, which is the higher limit, means "completely dependent, wishes to

TABLE 2

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF CLINIC AND SCHOOL CHILDREN ON ROGERS' TEST OF PERSONALITY ADJUSTMENT

	Clinic Boys ($N = 34$)		School Boys ($N = 49$)				Clinic Girls ($N = 4$)		School Girls ($N = 47$)			
	Mean	*SD*	Mean	*SD*	*t*	*p*	Mean	*SD*	Mean	*SD*	*U*	*p*
Personal Inferiority	14.35	4.46	9.79	3.54	5.19	.001	17.25	5.54	10.72	4.63	2.05	.02
Social Maladjustment	17.44	5.01	14.59	4.63	2.64	.05	17.25	4.15	15.81	4.15	.50	ns
Family Maladjustment	9.12	3.50	7.77	3.40	1.75	ns	11.75	6.42	6.13	3.66	1.56	.06
Daydreaming	2.94	2.22	1.73	1.93	2.62	.05	4.75	2.85	2.79	2.18	1.54	.06
Total Personality Adjustment	43.85	8.09	34.00	6.52	6.08	.001	51.00	8.51	35.94	7.35	2.68	.005

remain infantile." A score of 1 on daydreaming means no daydreaming at all, while a score of 4 means "extreme daydreamer, of fantastic types." Unfortunately the same two teachers did not evaluate all the children so that commonly accepted measures of reliability like Pearson's r could not be applied.

Originally the Rating Scale had been constructed to be used by child guidance clinic workers. A few changes were made to adapt it for school teachers so that the teacher could rate a child according to what the child did in school in relation to teachers and peers rather than what he did at home in relation to parents and siblings. The total adjustment rating comprises seven subscales: daydreaming, rebelliousness, inferiority, affection, independence, conflict over sex matters, and the child's general emotional adjustment. The last two subscales were not considered individually because many teachers, all women, failed to rate children on conflict over sex matters and because the individual subrating of emotional adjustment would have added little information by itself. Each subscale consisted of one question with a choice of four or five statements of increasing degree of maladjustment. The inferiority ratings, however, are a composite of two questions, hence the greater numerical value for this subscale (Table 4). All of the mean ratings reported are the sum of the ratings of both teachers for each subscale.

With two different teachers rating each child, the reliability of their ratings could only be estimated (Ebel's method, 1951). The teachers' ratings were on the whole surprisingly reliable (Table 3). With exceptions of ratings of inferiority in boys and daydreaming in girls, the high level of statistical significance reached by the ratings insures their usefulness as an external criterion for the CMAS.

TABLE 3

RELIABILITY FOR TEACHERS' RATINGS OF SCHOOL CHILDREN ON ROGERS' ADJUSTMENT RATING SCALE (Means and *SD*s in Table 4)

	Boys ($N = 49$)		Girls ($N = 47$)	
	F	*p*	*F*	*p*
Total Adjustment	2.88	.005	4.13	.001
Daydreaming	1.76	.05	.84	ns
Rebelliousness	3.38	.001	3.44	.001
Inferiority	.83	ns	2.64	.005
Affection	2.66	.005	1.23	ns
Independence	2.72	.005	2.79	.005

RESULTS AND DISCUSSION

Although the mean IQ difference between boys and girls was statistically significant ($t = 3.32$, $p < .01$), no significant correlation between manifest anxiety level and intelligence for either sex was found ($r = .00$ in both sexes). There were no differences between the means of boys and girls on anxiety level and on the lie score of the CMAS or on personal inferiority, social maladjustment, and total personality maladjustment as measured by Rogers' Test. The mean score for family maladjustment was slightly higher in boys than in girls, while girls were slightly higher than boys on daydreaming. According to their ratings, teachers viewed boys as being consistently more maladjusted than girls in all aspects of Rogers' Scale (Table 4).

Highly significant correlations between subtests and the total score of Rogers' Tests as well as between teachers' subratings and the total ratings of maladjustment in both sexes were found. The correlation between the total score for personality maladjustment on Rogers' Test and the total sum of ratings on Rogers' Scale was statistically significant for boys and nearly so for girls.

TABLE 4

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS FOR BOYS AND GIRLS ON THE CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE

	Boys ($N = 49$)		Girls ($N = 47$)			
	Mean	*SD*	Mean	*SD*	*t*	*p*
Anxiety	17.45	5.86	18.38	6.77	.72	ns
Lies	3.12	1.92	2.74	2.42	.85	ns
Personal Inferiority	9.79	3.54	10.72	4.63	1.10	ns
Social Maladjustment	14.59	4.63	15.81	4.15	1.36	ns
Family Maladjustment	7.77	3.40	6.13	3.66	2.27	.05
Daydreaming	1.73	1.93	2.79	2.18	2.52	.05
Total Personality Adjustment	34.00	6.52	35.94	7.35	1.37	ns
Total Adjustment Rating	30.92	6.42	25.42	6.09	4.29	.001
Daydreaming Rating	3.31	1.04	2.70	.83	3.18	.01
Rebelliousness Rating	4.43	.91	3.79	.58	4.13	.01
Inferiority Rating	7.73	2.82	6.25	2.12	2.88	.01
Affection Rating	3.47	1.19	2.83	.87	3.00	.01
Independence Rating	3.94	1.19	3.21	1.20	2.98	.01

Correlations between anxiety and all other scores and ratings for boys were low, statistically nonsignificant, and consistently negative. Boys who reported more feelings of inferiority on Rogers' Test were also rated as being more maladjusted, more dependent, and as possessing more feelings of inferiority than boys not reporting the same feelings. Those boys who reported greater family maladjustment were also rated as being more maladjusted, as daydreaming more frequently, as possessing more inferiority feelings, and as lacking in affection. Boys who reported more overall maladjustment were also rated as feeling inferior and as lacking in affection. The two latter ratings correlated also with ratings of rebelliousness (Table 5).

In girls two positive and significant correlations were found between anxiety and the total personality maladjustment score as measured by Rogers' Test as well as between anxiety and teachers' ratings of independence. The more anxious and maladjusted the girl, the greater was the probability of her being rated as being more dependent. Girls who reported greater family maladjustment were also rated as being more rebellious. Girls who reported greater personality maladjustments were rated as being more dependent. Self-reports of daydreaming correlated with ratings of greater rebelliousness, more intense feelings of inferiority, and more dependence. The latter ratings correlated with ratings of lack of affection and of greater feelings of inferiority (Table 6).

Although not all of the sex differences found were in the predicted direction, especially for boys, three additional analyses of the results were made to specify further the nature of such differences. In the first place, if manifest anxiety is related more to autoplastic rather than to alloplastic behavior, then girls should show a greater number of significant correlations between the measures of adjustment used. Girls do differ from boys on several correlations. A highly significant difference was found between correlations of anxiety and the total score on Rogers' Test ($z = 2.94$, $p < .01$). Girls showed significantly higher correlations than boys between anxiety and daydreaming scores ($z = 2.16$, $p < .05$); between ratings of maladjustment and of independence ($z = 2.27$, $p < .05$); and between ratings of inferiority and of independence ($z = 2.05$, $p < .05$). The only significant difference between correlations in which boys exhibited a higher correlation than girls was between ratings of rebelliousness and of lack of affection ($z = 2.25$, $p < .05$). The differences between correlations are in the expected direction, but could also be attributed to the teachers rating boys as being more maladjusted than girls.

In the second place, if girls do tend to use

TABLE 5

CORRELATIONS BETWEEN CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE FOR 49 BOYS

	2	3	4	5	6	7	8	9	10	11	12	13
1. Anxiety	.00	−.13	−.12	.04	−.20	−.19	−.07	.12	−.03	−.22	−.04	−.03
2. Lies		−.04	.09	−.17	.06	−.02	.16	−.04	.31*	.04	.15	.11
3. Personal Inferiority			−.23	.13	−.04	.42**	.42**	.24	.11	.42**	.19	.42**
4. Social Maladjustment				−.07	.09	.57**	−.06	−.04	−.06	.09	−.01	−.20
5. Family Maladjustment					−.01	.52**	.41	.31*	.04	.34*	.28*	.19
6. Daydreaming						.32*	.26	.00	.09	.18	.22	.24
7. Total Personality Adjustment							.46**	.26	.07	.51**	.30*	.25
8. Total Adjustment Rating								.46**	.52**	.82**	.66**	.59**
9. Daydreaming Rating									.27	.17	.10	.23
10. Rebelliousness Rating										.33*	.48**	.06
11. Inferiority Rating											.58**	.46**
12. Affection Rating												.06
13. Independence Rating												

* $p < .05$. ** $p < .01$.

TABLE 6

CORRELATIONS BETWEEN CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE FOR 47 GIRLS

	2	3	4	5	6	7	8	9	10	11	12	13
1. Anxiety	−.17	.11	.21	.24	.21	.40**	.26	.06	−.08	.24	.09	.30*
2. Lies		.08	−.13	−.02	.08	−.06	.05	.01	.16	−.03	.14	.03
3. Personal Inferiority			−.20	.30*	.04	.64**	.16	.11	.12	.10	.10	.23
4. Social Maladjustment				−.04	.13	.37**	.18	.11	−.22	.23	−.04	.24
5. Family Maladjustment					.07	.60**	.17	.28	.40**	.07	−.02	.17
6. Daydreaming						.41**	.02	−.20	−.03	−.06	.31*	.01
7. Total Personality Adjustment							.27	.16	.16	.20	.14	.34*
8. Total Adjustment Rating								.63**	.32*	.82**	.48**	.82**
9. Daydreaming Rating									.36*	.34*	.17	.46**
10. Rebelliousness Rating										04	.05	.22
11. Inferiority Rating											.38**	.73**
12. Affection Rating												.24
13. Independence Rating												

* $p < .05$. ** $p < .01$.

TABLE 7

ITEMS OF THE CHILDREN'S MANIFEST ANXIETY SCALE CHECKED POSITIVELY MORE OFTEN BY EITHER SEX

Item Number	Description of Item	x^2	p
	Girls More than Boys		
11.	I am secretly afraid of a lot of things.	4.13	.05
28.	My feelings get hurt easily.	9.60	.001
32.	It is hard for me to go to sleep at night.	4.38	.05
39.	I am afraid of the dark.	11.98	.001
	Boys More than Girls		
8.	I wish I could be very far away from here.	5.01	.025
10.	(Lie Scale) I would rather win than lose in a game.	10.09	.001
41.	(Lie Scale) I never get angry.	4.00	.05

autoplastic behavior more than boys, they should show it in the way they report about themselves. Indeed, an item analysis of the CMAS (Table 7) indicates that girls checked more often those items that show a greater tendency to admit feelings that boys would be leary of reporting, as Sarason and others have recently suggested (Sarnoff, Lighthall, Waite, Davidson, & Sarason, 1958).

In the third place, additional developmental trends may account for some of the sex differences. Boys showed variations in personal inferiority scores ($F = 2.59$, $p < .05$), with an increase up to sixth grade and a subsequent decline. Also among boys, daydreaming scores decreased significantly with age ($F = 3.16$, $p < .025$), reaching a mean of 1.00 in eighth grade. The decreases with age in daydreaming scores among girls was less pronounced ($F = 2.44$, $p < .10$). Variations in rebelliousness and daydreaming ratings with no clear declines or increases with age were also present among boys ($F = 3.04$, $p < .05$ and $F = 4.25$, $p < .01$, respectively). Girls showed variations in total adjustment ratings ($F = 2.64$, $p < .05$), daydreaming ratings ($F = 7.21$, $p < .001$), and ratings of affection ($F = 3.32$, $p < .01$), with no clear declines or increases with age. No other developmental trends were statistically significant in either sex.

In conclusion, although mean anxiety scores may be the same for boys and girls, yet each sex seems to cope with anxiety with different defenses. Perhaps the construction of separate anxiety scales for boys and girls should be considered.

SUMMARY

The present study evaluates some aspects of sex differences between boys and girls in their handling of anxiety. Although children may obtain equal scores on a paper and pencil measure of anxiety like the Children's Manifest Anxiety Scale, the relationships between manifest anxiety scores and various measures of adjustment may be different in boys and girls. It was expected that girls would show positive correlations between manifest anxiety level and measures of autoplastic adjustment, like daydreaming and feelings of inferiority. Boys were expected to show positive correlations between manifest anxiety level and alloplastic defenses, like rebelliousness and greater "nuisance value" than girls. To investigate whether any sex differences existed according to these expectations, the Children's Manifest Anxiety Scale was correlated with Rogers' Test of Personality Adjustment and Rogers' Rating Scale of Adjustment.

Boys were consistently rated by their teachers as being more maladjusted than girls, but no positive correlation between manifest anxiety scores and various measures of adjustment used was found for boys. Positive and statistically significant correlations between manifest anxiety level and Rogers' total score

of personality adjustment and ratings of independence were found for girls. Additionally, sex differences between correlations, in an item analysis of the Children's Manifest Anxiety Scale, and in developmental trends within each measure of adjustment were presented and discussed briefly.

REFERENCES

BONAPARTE, MARIE. *Female sexuality*. New York: Int. Univer. Press, 1953.

CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, **16**, 317–326.

EBEL, H. J. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 407–424.

FREUD, ANNA. *The ego and the mechanisms of defense*. New York: Int. Univer. Press, 1946.

GREENACRE, PHILLIS. Anatomical structure and superego development. *Amer. J. Orthopsychiat.*, 1948, **18**, 636–648.

KUHLEN, R. G. *The psychology of adolescent development*. New York: Harper, 1957.

LEVITT, E. E. Ecological differences in performance on the Children's Manifest Anxiety Scale. *Psychol. Rep.*, 1957, **3**, 281–286.

MCCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, **27**, 379–382.

ROGERS, C. R. *Manual of directions: A test of personality adjustment*. New York: Assn. Press, 1931. (a)

ROGERS, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contr. Educ.*, 1931, No. 458. (b)

SARNOFF, I., LIGHTHALL, F., WAITE, R., DAVIDSON, K., & SARASON, S. A cross-cultural study of anxiety among American and English school children. *J. educ. Psychol.*, 1958, **49**, 129–136.

SILVERBERG, W. V. *Childhood experience and personal destiny: A psychoanalytic theory of neurosis*. New York: Springer, 1952.

SMITH, L. M. The concurrent validity of six personality and adjustment tests for children. *Psychol. Monogr.*, 1957, **72**(4, Whole No. 457).

TERMAN, L. M., & TYLER, LEONA E. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954, Pp. 1064–1114.

WIRT, R. D., & BROEN, W. E., JR. The relation of the Children's Manifest Anxiety Scale to the concept of anxiety as used in the clinic. *J. consult. Psychol.*, 1956, **20**, 482.

(Received August 28, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 349-354

A NEW SCALE OF SOCIAL DESIRABILITY INDEPENDENT OF PSYCHOPATHOLOGY

DOUGLAS P. CROWNE

Ohio State University

AND DAVID MARLOWE

College of Medicine, University of Kentucky

It has long been recognized that personality test scores are influenced by non-test-relevant response determinants. Wiggins and Rumrill (1959) distinguish three approaches to this problem. Briefly, interest in the problem of response distortion has been concerned with attempts at statistical correction for "faking good" or "faking bad" (Meehl & Hathaway, 1946), the analysis of response sets (Cronbach, 1946, 1950), and ratings of the social desirability of personality test items (Edwards, 1957). A further distinction can be made, however, which results in a somewhat different division of approaches to the question of response distortion. Common to both the Meehl and Hathaway corrections for faking good and faking bad and Cronbach's notion of response sets is an interest in the *test behavior* of the subject (*S*). By social desirability, on the other hand, Edwards primarily means the "scale value for any personality statement such that the scale value indicates the position of the statement on the social desirability continuum . . ." (1957, p. 3). Social desirability, thus, has been used to refer to a characteristic of *test items*, i.e., their scale position on a social desirability scale.

Whether the test behavior of *S*s or the social desirability properties of items are the focus of interest, however, it now seems clear that underlying both these approaches is the concept of statistical deviance. In the construction of the MMPI *K* scale, for example, items were selected which differentiated between clinically normal persons producing abnormal test profiles and clinically abnormal individuals with abnormal test profiles, and between clinically abnormal persons with normal test profiles and abnormal *S*s whose test records were abnormal. Keyed responses to the *K* scale items tend to be statistically deviant in the parent populations. Similarly, the development of the Edwards Social Desirability Scale (SDS) illustrates this procedure. Items were drawn from various MMPI scales (*F*, *L*, *K*, and the Manifest Anxiety Scale [Taylor, 1953]) and submitted to judges who categorized them as either socially desirable or socially undesirable. Only items on which there was unanimous agreement among the 10 judges were included in the SDS. It seems clear that the items in Edwards SDS would, of necessity, have extreme social desirability scale positions or, in other words, be statistically deviant.

Some unfortunate consequences follow from the strict use of the statistical deviance model in the development of social desirability scales. With items drawn from the MMPI, it is apparent that in addition to their scalability for social desirability the items may also be characterized by their content which, in a general sense, has pathological implications. When a social desirability scale constructed according to this procedure is then applied to a college student population, the meaning of high social desirability scores is not at all clear. When *S*s given the Edwards SDS deny, for example, that their sleep is fitful and disturbed (Item 6) or that they worry quite a bit over possible misfortunes (Item 35), it cannot be determined whether these responses are attributable to social desirability or to a genuine absence of such symptoms. The probability of occurrence of the symptoms represented in MMPI items (and incorportated in the SDS)

in a college undergraduate population is undoubtedly low. Thus, the achievement of high SD scores may simply reflect the low frequency of pathological symptoms in this population and not the *needs* of *Ss* to present themselves in a favorable light. Of course, if one is only concerned with the properties of test items (their social desirability scalability), this is not a relevant issue. If, however, major importance is attached to the needs of *Ss* in psychometric situations and the influence of these needs on test responses, it is essential to be able to discriminate between the effects of item content and the needs of *Ss* to present themselves in a socially desirable (or undesirable) light.

In the present research, a social desirability scale was developed according to a different psychometric model, avoiding the ambiguities of the statistical deviance approach. Basic to this model is the sampling procedure employed in the selection of items from a defined universe. The population from which items were drawn is defined by behaviors which are culturally sanctioned and approved but which are improbable of occurrence. This will readily be recognized as the rationale underlying the Lie scale of the MMPI (Meehl & Hathaway, 1946). Items in the present scale, however, are probably less extreme than the Lie items.

Method

Development of Scale

A number of current personality inventories were consulted by the authors in order to devise a set of items for a new social desirability scale (M-C SDS). For inclusion in the scale, an item had to meet the criterion of cultural approval described above and was required to have minimal pathological or abnormal implications if responded to in either the socially desirable or undesirable directions. A set of 50 items meeting these criteria was submitted to 10 judges, both faculty members and graduate students in the Department of Psychology of Ohio State University, for social desirability ratings. The judges were instructed to score each item in the socially desirable direction from the point of view of college students, using true and false response categories. Unanimous agreement was obtained on 36 items and 90% agreement on 11 additional items. These 47 items constituted the preliminary form of the scale.

A major objective in the development of the M-C SDS was the elimination of pathology-relevant item content. To test this and for comparative purposes, both the M-C SDS and the Edwards 39-item SDS (Edwards, 1957) were submitted to 10 additional judges, again including both faculty members and graduate students in the psychology department, for ratings of the degree of maladjustment implied by socially undesirable responses to the items. A 5-point scale, ranging from extremely well-adjusted (1) to extremely maladjusted (5), was employed for this purpose. The mean rating for all the items in the M-C SDS was 2.8, slightly below the midpoint of the scale (implies neither good nor poor adjustment). The mean rating for the Edwards SDS items was 3.9, indicating that the judges considered socially undesirable responses on this scale to be definitely indicative of maladjustment. The t test of the significance of the difference between these means is 15.27, which is significant well beyond the .0001 level.

The preliminary scale was then administered to 76 students in two introductory psychology courses, and an item analysis completed. There were 33 items that discriminated at the .05 level or better between high and low total scores. Of the 33 items, 18 are keyed true and 15 false, making a response set interpretation of scores highly improbable. These 33 items constitute the final form of the M-C SDS and are listed in Table 1 with the socially desirable response scoring indicated.

Reliability

The internal consistency coefficient for the final form of the scale, using Kuder-Richardson formula 20, is .88. This was computed on 39 *Ss*, 10 males and 29 females, who were enrolled in an undergraduate abnormal psychology class at Ohio State University. The mean age of this sample was 24.4 years, with a range of 19 to 46 years. Thirty-one of these *Ss* took the scale on two occasions separated by a month interval. A test-retest correlation of .89 was obtained.

Relationship to Edwards SD Scale

The correlation between the M-C SDS and the Edwards SDS is .35, which is significant at the .01 level. The sample on which this correlation is based included, in addition to the 39 abnormal psychology students, 81 students in a course on exceptional children. The correlation shows a general tendency for scores on the two tests to be associated.

In Table 2, the means and standard deviations of both SDSs are reported. The distribution of M-C SDS scores rather closely approximates a normal distribution, while negative skewness, consistent with previous findings (Edwards, 1957), is found for the Edwards *SD* distribution. It is interesting to compare the Edwards SD mean found in the present research with that originally reported. The means of 28.6 and 27.1 for males and females reported by Edwards are considerably lower than the value found in this study.

Correlations with Other Scales

A considerable portion of the research on social desirability has involved the correlation of SDSs with MMPI variables. To compare the present scale with

TABLE 1

THE MARLOWE–CROWNE SOCIAL DESIRABILITY SCALE

Personal Reaction Inventory

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is *true* or *false* as it pertains to you personally.

1. Before voting I thoroughly investigate the qualifications of all the candidates. (T)
2. I never hesitate to go out of my way to help someone in trouble. (T)
3. It is sometimes hard for me to go on with my work if I am not encouraged. (F)
4. I have never intensely disliked anyone. (T)
5. On occasion I have had doubts about my ability to succeed in life. (F)
6. I sometimes feel resentful when I don't get my way. (F)
7. I am always careful about my manner of dress. (T)
8. My table manners at home are as good as when I eat out in a restaurant. (T)
9. If I could get into a movie without paying and be sure I was not seen I would probably do it. (F)
10. On a few occasions, I have given up doing something because I thought too little of my ability. (F)
11. I like to gossip at times. (F)
12. There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)
13. No matter who I'm talking to, I'm always a good listener. (T)
14. I can remember "playing sick" to get out of something. (F)
15. There have been occasions when I took advantage of someone. (F)
16. I'm always willing to admit it when I make a mistake. (T)
17. I always try to practice what I preach. (T)
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people. (T)
19. I sometimes try to get even rather than forgive and forget. (F)
20. When I don't know something I don't at all mind admitting it. (T)
21. I am always courteous, even to people who are disagreeable. (T)
22. At times I have really insisted on having things my own way. (F)
23. There have been occasions when I felt like smashing things. (F)
24. I would never think of letting someone else be punished for my wrongdoings. (T)
25. I never resent being asked to return a favor. (T)
26. I have never been irked when people expressed ideas very different from my own. (T)
27. I never make a long trip without checking the safety of my car. (T)
28. There have been times when I was quite jealous of the good fortune of others. (F)
29. I have almost never felt the urge to tell someone off. (T)
30. I am sometimes irritated by people who ask favors of me. (F)
31. I have never felt that I was punished without cause. (T)
32. I sometimes think when people have a mistortune they only got what they deserved. (F)
33. I have never deliberately said something that hurt someone's feelings. (T)

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE SOCIAL DESIRABILITY SCALES

Scale	*N*	Mean	*SD*
M-C SDS	120	13.72	5.78
Edwards SDS	120	31.83	5.06
(From Edwards, 1957)	84	28.6 Males	6.5
	108	27.1 Females	6.5

the Edwards SDS, Pearson product-moment correlations were computed between the two SDSs and the following MMPI and derived scales: *K*—Test-taking attitude; *L*—Lie; *F*—Validity and test-taking attitude; *Hs*—Hypochondriasis; *D*—Depression; *Hy*—Hysteria; *Pd*—Psychopathic Deviate; *Pa*—Paranoia; *Pt*—Psychasthenia; *Sc*—Schizophrenia; *Ma*—Manic; *Pr*—Prejudice (Gough, 1951); *St*—Status (Gough, 1948); *Es*—Ego Strength (Barron, 1953); MAS—Manifest Anxiety (Taylor, 1953); *A*—Anxiety (Welsh, 1956); *R*—Repression (Welsh, 1956).

The 39 *Ss* referred to above who served in the study were administered the M-C SDS, the 39-item Edwards SDS, and the MMPI in that order. The first two tests were given on the same day and the MMPI about a month later. Thirty-four *Ss* completed all of the tests and 37 of them completed all but the derived MMPI scales.

Table 3 presents the correlations between the M-C SDS and the Edwards SDS and the 17 MMPI validity, clinical, and derived scales. It is at once apparent that uniformly higher correlations obtain between the Edwards SDS and the various MMPI scales than between the M-C SDS and these MMPI variables. A general trend, which is consistent with previous research, is found in the positive correlations between the SDSs and the validity scales of the MMPI, and negative correlations with most of the clinical scales. Four clinical scales correlate highest with both SDSs, with the single exception of *D* which correlated −.27 with the M-C SDS: *Sc*, *Pd*, *Pt*, and *Hs*. Two of these four, *Sc* and *Pt*, are considered to be among the most "pathological" of the clinical scales.

TABLE 3

CORRELATIONS BETWEEN THE SOCIAL DESIRABILITY SCALES AND VARIOUS MMPI SCALES FOR 37 MALES AND FEMALES

MMPI Scales	M-C SDS	Edwards SDS
K	.40*	.65**
L	.54**	.22
F	−.36*	−.61**
Hs	−.30	−.62**
D	−.27	−.72**
Hy	.15	.09
Pd	−.41**	−.73**
Pa	.21	−.02
Pt	−.30	−.80**
Sc	−.40*	−.77**
Ma	−.24	−.42*
Pr[a]	−.27	−.58**
St[a]	.16	.14
Es	.17	.46**
MAS[b]	−.25	−.75**
A[b]	−.23	−.61**
R[b]	.28	.07

* Significant at the .05 level.
** Significant at the .01 level.
[a] $N = 36$.
[b] $N = 34$.

DISCUSSION

The most important feature of the findings of this study is found in the marked differences overall in the magnitude of the correlations between the two SDSs and the MMPI. Consistently higher correlations were found between the Edwards SDS and the MMPI scales than were obtained between the M-C SDS and the MMPI scales. The high Edwards SDS-MMPI correlations, in general, confirm findings previously reported by Edwards (1957) and Fordyce (1956). Correlations between the Edwards SDS and the *Pt*, *Sc*, and MAS scales, in fact, approach the asymptotic value of the reliabilities of the separate tests. With correlations this high, it is necessary to raise the question of whether the Edwards SDS and these MMPI scales are not, in effect, functionally unitary. It would appear to be difficult to hold the view that SD scores and MAS, *Pt*, and *Sc* scores can be interpreted differently. More in accord with the evidence would be to attribute the covariance of the Edwards SDS and these MMPI scales to item similarity and to the "pathological" content of both sets of items. This would lead to an interpretation of the Edwards scale as a measure of the willingness to admit to certain symptoms of a "neurotic" nature or as a measure of general "neuroticism." But this does not enable one to discriminate between high SD scorers who genuinely do not have the symptoms represented in the SDS items from those *Ss* who conceal (consciously or unconsciously) their symptoms and whose responses are motivated by social desirability. To the extent, then, that the Edwards SDS measures social desirability, it does so in the very re-

stricted sense that high *SD* scores imply that it is bad or undesirable to have or admit to symptoms. Possibly, such attitudes have little generality and would not be related to other test behavior or social behavior. Sarason (1959) has also raised the question of the interpretation of Edwards SDS as an unconfounded measure of social desirability.

In the development of the M-C SDS, social desirability was defined more broadly to refer to the need of *S*s to obtain approval by responding in a culturally appropriate and acceptable manner. This conception does not involve the acquiescence or denial by *S* of pathology. The significantly different maladjustment ratings obtained on the two SDSs support the hypothesis that the Edwards SDS involves the admission or denial of maladjustive symptoms and indicate that socially undesirable responses on the M-C SDS do not imply maladjustment.

The smaller correlations between the M-C SDS and the various MMPI scales would be predicted if one views social desirability as accounting for a fraction of the MMPI variance but not all or most of it. The problem of overlapping meanings is thereby avoided. Thus, it is submitted that the M-C SDS-MMPI correlations more accurately indicate the amount of MMPI scale variance which may be attributed to differences in the need to give socially desirable responses.

It may additionally be pointed out that the M-C SDS and the Edwards SDS differ considerably in the amount of content or item overlap with the various MMPI scales. The present scale contains one exact and four approximate replications of *L* items and one repetition of a *K* scale item. By contrast, the Edwards SDS, it will be recalled, was constructed from a heterogeneous pool of MMPI items and not inconsiderably overlaps with many MMPI validity and clinical scales. The two SDSs have no items in common.

Certain additional aspects of the present findings are worthy of note. Positive correlations are found for both SDSs with the *K* and *L* scales, on which high scores are generally interpreted to indicate "defensiveness" and the attempt by *S* to cast himself in a favorable light. The M-C SDS correlates much more highly with *L*, however, than does the Edwards SDS. The negative correlations with the *F* scale are accounted for by the interpretation of a high *F* score as an indication of "plus getting." Regarding the clinical and derived scales, in general those MMPI scales on which a high score indicates maladjustment are negatively correlated with the SDSs. In part, the exceptions to this may be explained in terms of the distinction between subtle and obvious scoring on some of the clinical scales. Item subtlety, meaning the relative absence of social desirability implications, would account for the negligible correlations between the Edwards SDS and the *Hy* and *Pa* scales, for example. The fairly substantial correlations between the Edwards SDS and the *Pr* and *Es* scales may again be a function of similarity in general item content. In the judgment of the present authors, about half of the items in the *Es* scale would be classed as "pathological," while roughly a third of the *Pr* items would be so considered.

The positive correlation between the M-C SDS and the *Pa* scale, however, is an interesting possible exception. While this *r* falls short of significance, it might suggest that high SDS scores (implying in the present definition of the construct a high need for the approval of others) tend to be associated with concern or suspicion about the motives of others. Correlations between the Edwards SDS and Welsh's *A* and *R* scales have not, to the writers' knowledge, been previously reported. The Edwards SDS correlated highly, as one would expect, with the *A* scale but not at all with *R* which has a rather heterogeneous item content in terms of pathology. The M-C SDS does better in this case with an *r* of .28. This is in the predicted direction since all of the items on the *R* scale are keyed false. The M-C SDS correlation with the *A* scale is of the same magnitude as the correlation with the MAS and is consistent with expectation. It would appear from the correlations of the SDSs with *A* and MAS that the latter are approximately equivalent measures.

Summary

In this research, an alternative model to Edwards' conception of social desirability was proposed. Basic to the present construct of

social desirability is the definition of a population of culturally acceptable and approved behaviors which are, at the same time, relatively unlikely to occur. Test items were drawn from this population in the development of a new social desirability scale, the Marlowe-Crowne Social Desirability Scale. This scale was correlated with 17 MMPI validity, clinical, and derived scales and the results compared with the correlations of the Edwards SDS with these MMPI variables. The very high correlations obtained with the Edwards scale cast doubt on the interpretation of this test as a measure of the influence of social desirability on test responses. The magnitude of the correlations of the new scale with the MMPI was considered to be more in accord with a definition of social desirability in terms of the need of subjects to respond in culturally sanctioned ways.

REFERENCES

BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 475–494.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

GOUGH, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.

GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-semitism. *J. soc. Psychol.*, 1951, **33**, 247–255.

MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, **30**, 525–564.

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *J. abnorm. soc. Psychol.*, 1959, **59**, 37–42.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

WELSH, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R*. *J. consult. Psychol.*, 1959, **23**, 100–106.

(Received August 29, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 355–360

RELATIONSHIP OF FAMILY DISTURBANCE TO COGNITIVE DIFFICULTIES IN A LEARNING-PROBLEM CHILD[1]

MICHAEL A. WALLACH,
Massachusetts Institute of Technology
DAVID N. ULRICH,
Darien Public Schools, Connecticut
AND MARGARET B. GRUNEBAUM
Judge Baker Guidance Center, Boston, Massachusetts

Psychologists interested in the relationship of motivation to cognition have obtained some understanding of the ways in which cognitive impairment can result from emotional disturbance. Meanwhile, clinical inquiries have led to an expanding knowledge of the effects of parent-child relationships upon the emotional processes of the individual. Little has yet been done, however, to trace the sequence from family relationships, through the child's emotional dynamics, to his cognitive processes. The present paper is designed to help bridge these areas by tracing a causal linkage between disturbance in parent-child interactions and the child's cognitive impairment.

The boy in question, Robert Strong, is now 15. He has had weekly hours of therapy for four years and has seen a therapeutically-oriented tutor for one to three hours a week over two academic years. His mother has seen a social worker for weekly hours during this whole period, and his father has seen the same person for the past two years. Eleven when he first came to the clinic, Bobby had repeated the second grade and had had near-failing marks in each grade through the fifth. Yet at the time of referral the Wechsler-Bellevue Intelligence Scale for Children confirmed that his intelligence was average. After three and one-half years of treatment he has made marked gains in areas other than school performance, but he has yet to obtain an average report card. Regularly administered school achievement tests have left no doubt of his trouble in acquiring the basic skills and absorbing knowledge. Although these tests indicate that he is now above grade level in reading skill, this gain has not been reflected so far in his school performance.

[1] This research was supported in part by two grants from the United States Public Health Service. One of these is under the direction of Bessie M. Sperry, and the other, of Jerome S. Bruner, whom the authors wish to thank for encouragement and aid. We are indebted to Beatrice Reiner, whose two years of work with the mother contributed much to the case.

The Parental Relationship

The immediate family consists of Mr. and Mrs. Strong, a couple in their early 40's, their daughter, Flora, who attends college and comes home only on weekends, and Robert. Mr. Strong works as a semiskilled laborer in a factory, and his wife teaches nursery school.

On the surface, Mrs. Strong would wish us to see her as calm. She seeks to control the expression of emotions by reason and by rigid rules for behavior taken mainly from her religion. Education and the development of reasoning powers have been of great importance to her, since they help protect her against emotions and also constitute a means of expressing her rivalry with men. It is not a coincidence that her husband, in turn, is a person who has tiptoed through life. At 43 he is boyish in appearance and mannerism. In the beginning of treatment he described himself as "meek and shy, a self-conscious man with rusty brains." He bows out completely from the intellectual realm, deferring all to his

wife and stating that his only contribution is that of brawn. Like his wife, his major defense is that of denial. Were we to take his word, anger, resentment, and rivalry are nonexistent in his makeup. Yet, in truth, they accumulate to such an extent that he has explosive outbursts which frighten him as well as the rest of the family. In contrast to his wife's position, Mr. Strong has looked upon education and the ability to reason as powers which he could not afford to possess lest they make him destructive.

In the parents' marriage, a balance was formed that met both their needs. By virtue of her husband's need for her, Mrs. Strong has been able to deny her rivalry with men, to maintain at the same time a superior position to men, and also to continue to see herself as a helping person. If he feels unable to carve the meat, answer the phone, do the family finances, then she must take over these duties to help him. For Mr. Strong, the choice of marital partner was part of a lifelong renunciation.

Consider now the children of this marriage. Flora, their first-born, has managed to grow without disturbing this balance. She achieved a good academic standing and thereby formed a team with her mother of the educated women in the family. The role that was assigned to Bobby, on the other hand, is best understood by considering what the birth of a male child meant to these parents. For quite different reasons, a son was a threat to both of them. To Mrs. Strong, a male child was a potential competitor, a potential danger if he grew to mature manhood. To Mr. Strong, Bobby was perceived as a sibling rival. It was as though the father feared that Mrs. Strong would not have enough to give to two. At the same time, he saw that Bobby could possibly excel over him in manliness. We can see that the role Bobby adopted—that of passivity, dependence, and underachievement—met the needs of both his parents in quieting their fears.

The Boy's Emotional Difficulties

The parents' everyday behavior continually suggested that people contain dangerous impulses and the world is a dangerous place in which to live. Bobby had listened to his father's detailed accounts of crippling accidents at the factory. He had heard his mother's equally thorough descriptions of the motorcycle accidents that had killed his maternal grandfather and maimed his maternal grandmother long before Bobby was born. He had overheard, while lying frightened and alone in his crib, the sounds that accompanied his maternal grandmother's death from a heart attack and fall down the stairs. Events such as these could well reinforce fear and guilt about his own aggressive impulses, especially when his mother did not let him forget that the bigger he got, the more easily he could hurt someone. Bobby found his attempts at mastery were often received by his parents as a hostile effort to put someone else out of first place. School was actually the most important sphere of achievement in Bobby's life, and both his parents felt strong ambivalence about his success in this area. Accordingly, accomplishment in school became endowed for him with fear and guilt.

On one occasion when the therapist discussed Bobby's fears about achievement in school, Bobby made no verbal response but worked his hand across the desk in a walking movement. When asked about this, he said it reminded him of a man with a broken leg on crutches. The weekend after receiving his first really encouraging school report, Bobby broke his skis in a fall. When asked how his sister would like it if he got a good report card, he replied that she would have a fit. When asked how his tutor would like it, he replied that his tutor would hit the ceiling. Bobby did not realize until it was pointed out to him that these phrases implied any hostility. On another occasion, he brought in a test paper containing an A and absent-mindedly erased the A from the paper while talking about it.

Even when his daily work had improved, he continued getting poor report cards. The low grades could be attributed to various self-defeating maneuvers that will be discusssed below. No matter how genuine his distress about failure, it was evident that Bobby possessed a clearly defined need to fail.

Bobby seemed to perceive his relationship to his mother and father as laden with conflict. It was as if Bobby and the rest of the family were continually engaged in playing

"king of the mountain," and one could get enough space only by shoving someone else out of it. When asked if his father went ice skating with him, he replied with a grin, "No, I was using his ice skates." When Bobby proudly appeared in a pair of size 9 shoes, the therapist asked him what his father had said about this new possession. Bobby promptly answered, "He said now he could wear my shoes." When his wishes and fears about displacing father from the house were interpreted to Bobby, he replied with a dreamy smile, "You take the living room and I'll take the kitchen . . . and you'll starve to death."

Denial seemed to be Bobby's main means of reacting to family conflicts. He could never discuss clearly any central aspect of his relationship to his parents. When asked what was going on at home, he would say only that father had been doing some carpentry, or they had been working in the garden. During a period when the parents were known to be going through an explosive upheaval in their relation to each other, even the most persistent prodding brought no comment from Bobby. After extensive work with him, it began to appear that a close parallel existed between the way he thought about his family and the way he thought about schoolwork. Rather than grasp the crucial variables of any problem in their important relationships to one another, for instance, Bobby would fasten upon a fragmentary assortment of often peripheral details.

The Boy's Cognitive Processes

Schoolwork's Danger Potential

Bobby was highly sensitized to the danger potential of everything in his environment. No matter with what he was engaged, those attributes of the object or activity that might injure him were the ones of which he was most aware, and hence the ones to which he responded. During one session, for example, the tutor suggested a word game in which one person provides a key word, and then the other must write out, as quickly as possible, all the words that the key word suggests to him. When the tutor gave Bobby "danger" as the key word, he replied, "I can answer that one in a single word," and wrote "everything." He then proceeded to describe a car as something that runs you over, a plane as something that crashes, machines in factories as things into which the unwary worker can fall and get chopped up, a door as something that can catch and squash your finger, and on without end. In the case of each such object, attributes which are not very important to most children were the critical ones for Bobby and served to control his thought about the object.

If Bobby viewed everyday objects like cars and doors as potential sources of harm, this was even more true of his feelings toward school and tutorial sessions. Once, for example, when tutor and tutee were alternately making up radio "commercials" and recording them on tape, Bobby gave a commercial for the "Boston Singing Club," telling how you can get singing lessons, and as a result you develop a pitch that will break a glass. Singing lessons, his metaphor here for the tutorial sessions, apparently were seen as making him capable of doing harm.

Consider also the time when he and his tutor were drawing a graph to represent the relation of temperature to pressure of a certain gas. After working on the graph for a while, Bobby announced that he was sure there was going to be an explosion of the gas due to the pressure, so the pressure would go down; and he promptly extrapolated a dotted line on the graph which veered sharply downward. The tutor assured him that the scientist had designed a very strong steel chamber for the gas so it could not possibly explode even if the pressure were increased; but Bobby claimed rather that the chamber was bound to explode, the scientist get hurt, and the gas pressure drop. He then sketched in further drastic rises and falls of the dotted graph line, put snowcaps on all the peaks of the graphs—saying they are snow-covered mountains—and drew a figure skiing down the mountain side. Immediately afterward he drew a tree squarely in front of the skier, so the latter had to crash into it. When the tutor noted that a crash was imminent, Bobby then drew in two ski tracks going around opposite sides of the tree so that the skier would have to split in two in order to follow them. Thus, if Bobby could not claim his schoolwork would lead to

death through exploding gases, he would make it lead to death by some other means. Bobby's mind seemed far indeed from the ostensible goal of working on the graph. He was totally occupied with assaying the injury potential of his paper-and-pencil activity.

Much of Bobby's cognitive energy was absorbed by the task of detecting and guarding himself against such imagined dangers as these. He showed an avoidance of academic material in general and also an avoidance of the concept of equivalence in particular. We shall consider each in turn.

Physical Avoidance of Schoolwork

There were two levels at which Bobby sought to avoid academic material in general: the level of physical activity and the level of thinking. First of all, Bobby had developed to a high degree strategies for physically getting himself out of, or not getting into, schoolwork situations. Several times, for instance, he managed to "forget" to bring his books to tutorial, too often for it to be mere chance and despite specific injunctions not to forget. If the books were brought, another kind of procrastination often ensued: hiding so the tutor could not find him. When finally engaged in some sort of schoolwork, he would be as inactive at it as possible—yawning and sprawling over the table as if he had not slept for days, acting as if the pencil were too heavy a weight to lift. And at every possible juncture, he would try to lure the tutor away from the topic of schoolwork by starting to talk about other things or looking around for any object that might provide a distraction. At school he sometimes would do the wrong homework or fail to hand it in.

All children, to be sure, would rather do other things than schoolwork. But most of them learn to engage in school activities nevertheless, and even to find pleasure in them. Bobby, on the other hand, showed a degree of insistence in avoiding academic activities which went far beyond playful truancy.

Avoidance of Abstraction and Anticipation in Thinking

When physical means failed, Bobby's avoidance of academic material went on at the level of thinking itself. We have already suggested that Bobby strongly denied his knowledge of crucial family relationships. Confronted in school with the task of determining which were the important facts and how they were related, Bobby was often at a loss. It seemed as if he were trying to avoid the kind of thinking that might lead back to intolerable discoveries. One part of thinking has to do with *abstraction* activities: summarizing a potpourri of data in some terse and effective manner that is easier to retain and use than were the original facts. Such is the function of a formula or of an outline for a book. Another process in thinking concerns what we may call *anticipation* activities: predicting the consequences of present events by a kind of trial-and-error in imagination. The effect of avoiding such abstraction and anticipation activities will appear as a kind of "perceptual seduction" by the specific information made available to the senses.

Bobby, in his tutorial work, gave evidence of avoiding both aspects of thought. Consider first the avoidance of abstraction. With regard to verbal material, for instance, he often complained that he never knew what he should remember, in history or science, from the jumble of words before him. In his note-taking he would copy out nonessential details verbatim from the text, and often miss the whole point of a passage. His own writing, in turn, was filled with kinds of errors that indicated an inability to rise above specific words and take in the breadth of a whole sentence or thought. The particular words themselves, each separated from its neighbors, held his complete attention—as, for instance, in the following sentences he wrote on diabetes: "Diabeles is inabiliity of the bady ta oxidize sugar is known as diabetes. diabetes is a desiase in the body that insulen is scarce that causes diabites." This statement was written at age 13.

Avoidance of reasonable anticipations in his cognitive work was just as clear as his avoidance of abstraction. When assembling a radio kit in tutorial one day, he came to a place where three items had to be secured with a single screw. Bobby screwed one part in, discovered a minute later that another piece had to go there too and redid the whole screw assembly, then discovered soon again that yet a third piece was required and had to undo

everything once more. Later he came to another identical place and went through this same procedure, and still later this happened again. Even though the three assemblies were exactly the same, he did not profit from his past experience and anticipate that several parts had to be held by the same screw when he came to each subsequent screw assembly. Knowledge was never transferred, as if the planning that would be involved was a burden he refused to assume.

Avoidance of Equivalence Relationships

Bobby's avoidance of the threatening topic of relationships among family members spread to the matter of relationships among numbers too. Bobby had much difficulty with the notion that various arithmetic expressions are equivalent to one another—something his arithmetic teachers had been demonstrating to him for some time now. Even after studying these equivalences for weeks, he still could not tell the tutor the correct decimal equivalent of the fraction ½ or the percentage equivalent of that fraction. When forced to try an answer, he would guess at random or would respond in terms of the specifics of the ½ by claiming, for example, that ½% was the required percentage equivalent. In similar fashion, Bobby could not translate from one fraction to another. For example, he could not guess what simpler fraction would be the approximate equivalent of 250/525. Again, after tutor and tutee had indicated on a line the locations of 0, ¼, ½, and 1, Bobby could not tell where ⅜ would be located on this line, even though ¼ appeared there already; and he could not find the position of ⅔, even though ½ already appeared. This boy was 13 years old at the time.

During one session, Bobby was marking off sixths at equal intervals by means of vertical lines. He finally got to ⅚ and then went over one further equal interval and labeled it "1." He thus put down ⅚ and 1 at different positions and was quite satisfied with this until the tutor questioned him about it. It was as if he feared having the ⅚ and the 1 displace each other—and the only way he could avoid this was to imply that they were not equivalent but different. This latent fear could well derive from his conceiving of family relationships as involving no more than one possible actor for each role. When he was forced to deal actively with number equivalences, in turn, we found that once again he did not believe that two terms could fulfill a common function. His modes of thinking about the family hence may have generalized to an abstract cognitive domain.

What happened when Bobby was forced explicitly to consider number relationships is illustrated by such episodes as the following. The tutor, for example, set up the equation, ⅗ = .6, and asked Bobby: "Does this mean that the ⅗ gets lost somehow or destroyed when we put down the .6, or rather does it mean the .6 and ⅗ are just different ways of talking about the same thing?" Bobby decidedly indicated than an arithmetic term got lost, destroyed, when you made something else equal to it, and held firmly to this view of the matter despite lengthy explanations by the tutor on many different occasions. The tutor, for instance, presented such examples as ⅗ = ⅙₀, and had Bobby cancel through the ⅙₀ and thereby reduce it back to ⅗. Here too he thought the ⅗ got destroyed when made equal to the ⅙₀, and he did not feel his logic undermined when he found the ⅗ recoverable again from the ⅙₀. Even when he was finally able to grasp the idea in the case of one example, he was still quite incredulous when the tutor told him that this is also what went on in the other cases that were discussed.

Later the tutor demonstrated in yet another way that these two fractions really referred to the same thing, by illustrating each of them on a pie-cut circle graph of the same diameter: the segments representing ⅗ and ⅙₀ of each circle, hence looking exactly the same. Finding out in this way that the proportion was the same on both circle graphs appeared to cause Bobby the greatest surprise, and further demonstration of other arithmetic equivalences by the same method continually brought forth exclamations of amazement. All this at an age, 13, when work with arithmetic equivalences is second nature to most children.

When the tutor drove Bobby home after a session in which the meaning of arithmetic equivalence had been explored, Bobby was inordinately prone to see the environment as

injurious: expostulating at great length on how cars might smash up against the curb and hurt it, on how dangerous cars were in general, and then himself almost running into the path of an oncoming car while crossing the street. During another session on number equivalences, he squirmed, was uneasy, and eventually broke off from his work to describe newspaper accounts of a boy who got twisted up in a clothesline and hung himself by accident, and a girl who was climbing a tree when she fell into a Y formed by the branches and choked to death. During the work itself, Bobby had been punching holes viciously with a pencil into the side of a cardboard box.

With regard to numerical equivalence relationships, then, Bobby suffered an impairment in his capacity for what Piaget (1954) calls "conservation of identity." In these number realms, any different manifestation or appearance of something meant, for Bobby, a different thing, and could never mean the same thing. This trouble with the conservation of identity perhaps had its roots in his fear that equivalence in the family was impossible, that one of the aspirants had to suffer destruction.

This section has described the impact of Bobby's emotional difficulties on his cognitive activities. We have noted his sensitization to the possibility of danger in school activities, his physical and cognitive avoidance of school-related matters in general, and his avoidance of the concept of equivalence in particular. Our discussion has traced these cognitive impairments to two main kinds of causes: fear of school achievement and avoidance of knowledge of crucial family relationships.

Conclusion

We have sought to demonstrate how a long-standing disturbance in family relationships can result in cognitive difficulties for a child. The effort to avoid awareness of the family conflict has led to chronic interference with formal thought processes, while the meanings attached to school achievement have had a further inhibitory effect. Treatment over a long period has brought about a variety of marked changes in the parents and in the boy, and he seems to be meeting the other challenges of adolescence with a degree of success that could not have been predicted as recently as a year ago. In particular, he has overcome an initial strong shyness with respect to girls and now boasts a steady girl friend, he does very well in school sports and shop, and he has been making more friends among peers. However, his general thinking capacity still shows partial impairment, and his capacity for reproducing knowledge under the stress of a school setting is still markedly impaired. In retrospect, it appears that successful therapeutic intervention in a case of this kind requires a much closer collaboration of therapeutic and educational methods than has yet generally been achieved. Most instructive for psychology, perhaps, is the degree of both specific and general relationship found between family setting and the child's cognitive processes.

Reference

Piaget, J. *The child's construction of reality*. New York: Basic Books, 1954.

(Received September 2, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 361–367

CASE HISTORY CORRELATES OF THREE MMPI PROFILE TYPES[1]

HAROLD GILBERSTADT AND JAN DUKER

University of Minnesota

In spite of the accumulation of research studies attesting to the superiority of the actuarial over the clinical approach in describing personality (Meehl, 1954), surprisingly little attention has been given to the task of compiling descriptions to be applied actuarially. As a result of his survey of this research, Meehl (1956) called for the development of "cookbook" descriptions. The purpose of the present paper is to report on a project designed to develop a source book of personality descriptions for narrowly specified MMPI profiles. It is hoped that the use of such descriptions will improve communication about patients, will make profile interpretation less dependent on the personal experience and skill of the individual clinician, and will help to standardize classification of personality for further research. The present paper reports data from a trait and symptom check list which was applied to the case records of three MMPI types.

METHOD

Population

Records were selected from the files for all patients admitted to the Psychiatric Service of the Minneapolis Veterans Administration Hospital and examined by the psychology section during the years 1952–1957. Of approximately 3,150 such cases, about 2,300 were first admission. All patients were male veterans. First admission cases which met the following criteria were studied:

1. MMPI administered not more than 21 days before or 21 days after admission.
2. Age range 20 to 60 years.
3. Primary diagnosis not brain damage.
4. MMPI scale T scores: $L = 60$ or less, $F = 85$ or less, $K = 70$ or less.
5. Shipley Institute of Living Scale IQ estimate 105 or more.

Selection of Profile Types

Of several methods of profile analysis (Cronbach & Gleser, 1953; Lykken, 1956; Mosel & Roberts, 1954), it was decided that the contingency method employed by Meehl and Dahlstrom (Meehl, 1959) was best suited for the project. This method consisted of the empirical derivation of a set of rules which would specify the elevations and the interrelationships of the MMPI scales of a given type of profile and which would exclude profiles assumed to be not of the given type. The method permitted utilization of experience which the investigators had accumulated with respect to the MMPI profiles of the population being studied. On the basis of this experience, a set of specimen MMPI profiles was collected for study which was expected to be representative of frequently recurring personality types in the setting. Characteristics of three profile types will be reported in this paper; namely, MMPI codes 2–7–8 (*D-Pt-Sc*), 2–7 (*D-Pt*), and 2–7–4 (*D-Pt-Pd*). Starting from the specimen profiles, a set of con-

[1] The authors wish to express appreciation to Starke R. Hathaway and to Robert D. Wirt of the University of Minnesota for their helpful suggestions concerning the preparation of this paper.

TABLE 1

MEAN MMPI T SCORES FOR THE THREE CODE TYPES

	Scale													
	N	*L*	*F*	*K*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*	*Si*
Code type														
2-7-8	22	48	72	49	73	99	74	77	70	70	98	99	53	77
2-7	13	50	55	55	68	94	74	62	58	60	84	62	47	58
2-7-4	27	48	62	50	66	90	71	83	58	64	83	68	60	60

TABLE 2

PERCENTAGE FREQUENCIES OF CHECKLIST ITEMS FOR GENERAL ABNORMAL (GA) SAMPLE AND EACH PROFILE TYPE

		Greater than GA			Equal to or less than GA		
Check List Item	GA	2-7-8	2-7	2-7-4	2-7-8	2-7	2-7-4
Agitated	14		23		5		7
Anxiety	33	46	85**	56*			
Anxiety reaction diagnosis	25		76**	44	18		
Auditory hallucinations	4	14				0	4
Back pain	5		8	11	5		
Blindness, eye complaint	3	5	15				0
Blunted or inappropriate affect	18	41*				0**	0**
Cardiac complaint	3	5	23*	4			
Chest pain	3		15	4	0		
Compulsive	6	14	23	7			
Conflict with parents	14				5	0*	0**
Conflict with siblings	11				0**	0*	0**
Conflict with wife	30			56*	14	15	
Confusion (Nonorganic)	11				9	0*	0**
Dependent	19		23	30	18		
Depression	43	87**	85**	52			
Depressive reaction diagnosis	8	18	23	30**			
Difficult concentration	2	28**	8				0
Divorced	12			41**	5	0*	
Dizziness	6		15		5		0*
Epigastric complaint	17				9	15	11
Evasive-defensive	22				0**	8	4**
Father alcoholic	9				5	0*	7
Father domineering	4	23*				0	0
Father strict	3	18*	8	4			
Fearful	10	32*				8	7
Feelings of hostility	17				9	8	0**
Feelings of sex inadequacy	8	14				0	4
Financial difficulty	6		8	26**	5		
Guilt	13	18	15	26			
Headache	12	18	30				7
Heavy drinking	32			96**	5**	0**	
Hostile	10		23	15	9		
Ideas of reference	11	32*				0*	0**
Immature	17	28		30		8	
Impulsive	6			11	0*	0	
Inadequacy feelings	8	37**	23	11			
Indecision	3	23**		7		0	
Inferiority feelings	5	37**	23				4
Insomnia	11	14	46**	30*			
Irritable	11		30	26	9		
Loss of appetite	7	18	38*	15			
Loss of interest	2	23**		4		0	
Married	58		100**		14**		41
Mother physically ill	4	14				0	0
Neck pains	3	5	15	7			
Nervousness	30	37	69*	48			
Numbness	6	14	8				0*
Obsessions	1	18**	30**				0
Paranoid delusions	10				0**	0*	0**
Paranoid trends	10	14				8	4

TABLE 2 (*Continued*)

Check List Item	GA	Greater than GA			Equal to or less than GA		
		2-7-8	2-7	2-7-4	2-7-8	2-7	2-7-4
Passive	8	18				8	4
Poor work adjustment	43			45	23	23	
Restless	13		23		5		11
Rigid	3		15		0		0
Rumination	1	28**	15*	4			
Schizoid	8	46**				8	4
Schizophrenic reaction diagnosis	16	50**				0**	0**
Suicide attempt	6	9		15		0	
Suicidal preoccupation	12	23				0*	11
Suspiciousness	9				9	0*	0**
Tearfulness	7	14	8				4
Tension	21	41	69**	52**			
Tremor	14			26	14	0*	
Unmarried	30	82**				0**	7**
Weak, tired, fatigued	9	32*	38*				7
Weight loss	6	14				0	4
Wife pregnant or post	5		23	15	0		
Withdrawn, introversive	16	51**				15	4
Worrying	9		46**	19	9		

* Difference significant at .05 level.
** Difference significant at .01 level.

tingency rules was formulated for each type which would narrowly specify the characteristics of the profile type.

The case records for each of the three types were studied in the following manner. First, all cases for the year 1956 whose MMPI profile fit the rules and the criteria outlined above were selected. These records were read. They consisted of psychiatric discharge summaries and social case history data. This first reading gave a subjective clinical impression of the patients in the group. These records were then re-read and placed on a five-point scale from poor to excellent for goodness of fit to the preliminary subjective clinical formulation. Rules were then revised to eliminate cases which did not seem to fit well. Next, all records from the additional years of 1952 and 1953 were read and rules were again refined to eliminate cases which seemed inappropriate. Finally, the records for the remaining three years (1954, 1955, and 1957) were read and cases which did not receive high ratings for the developing types were considered to be misses. The cases which were eliminated in the process of refining the rules are not reported here in detail, although some characteristics of them will be discussed below.

The checklist was based on one developed empirically by Cantor (1952) who tallied descriptive terms appearing in discharge summaries of psychiatric patients in the Minneapolis Veteran's Administration Hospital. For the present study, three judges independently checked the list for each case. When two of the three judges agreed, the item was included in the frequency counts.

Results

Table 1 gives the mean *T* scores of the MMPI scales for each of the profile types.

Table 2 presents the incidence of the checklist items for the three code types and for patients in general (Rosen, 1952). Items which were of less than 10% frequency for at least one of the groups have been omitted from the table. Significances of the differences in item frequencies between the groups were computed from a nomograph for calculating the significance of the difference between two percentages (Lawshech and Baker, 1950).

2–7–8 Type ($N = 22$)

The following set of rules was developed to specify the 2–7–8 MMPI code type:

1. *D*, *Pt*, and *Sc* over *T* score 70.
2. *D* minus *Sc* less than 15 *T* scores.
3. *Pt* and *Sc* less than 20 *T* scores apart (unless *D* or *Pt* or *Sc* greater than 100 in which case discrepancy can be greater).
4. If *Pt* is peak, include only if all other scales are below 90 and *Pt* is not more than 5 *T* scores greater than *D*.
5. *Si* greater than *Ma*.
6. *Ma* below *T* score 70.
7. *D* minus *Hs* more than 10 *T* scores.

TABLE 3

COMPARISONS BETWEEN PROFILE TYPES ON CHECKLIST ITEMS

Check List Item	278 > 27	278 > 274	27 > 278	27 > 274	274 > 278	274 > 27
Anxiety			*			
Auditory hallucinations	*					
Blindness, eye complaint				*		
Blunted or inappropriate affect	**	**				
Chest Pain			*			
Conflict with wife					**	**
Depression		**		*		
Difficult concentration		**				
Dizziness				*		
Father domineering	**	**				
Fearful		*				
Feeling of sex inadequacy	*					
Financial difficulty					*	
Heavy drinking					**	**
Ideas of reference	**	**				
Impulsive					*	*
Inadequacy feelings		*				
Indecision	**					
Inferiority feelings		**				
Insomnia			*			
Loss of interest	**					
Mother physically ill	*	**				
Numbness		**				
Obsessions		**		**		
Poor Work Adjustment						*
Quiet	*	**				
Rigid			*	*		
Rumination		*				

* Difference significant at .05 level.
** Difference significant at .01 level.

TABLE 3 (*Continued*)

Check List Item	278 > 27	278 > 274	27 > 278	27 > 274	274 > 278	274 > 27
Schizoid	**	**				
Sexual difficulty	**					
Suicide attempt						*
Suicidal preoccupation	**					*
Suspiciousness		*				
Tremor	*					*
Weak-tired-fatigued		*		*		
Weight loss	*					
Wife pregnant or post			*		**	
Withdrawn-introversive	*	**				
Worrying			*			

8. *Pt* minus *Pd* more than 10 *T* scores.
9. *Pt* minus *Pa* more than 10 *T* scores.

Patients obtaining profiles of the 2–7–8 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: depression, difficult concentration, inadequacy feelings, indecision, inferiority feelings, loss of interest, obsessions, rumination, schizoid, schizophrenic reaction diagnosis, unmarried, weak-tired-fatigued, and withdrawn-introversive. Significantly different in frequency from patients-in-general at the .05 level were: blunted or inappropriate affect, father strict, fearful, and ideas of reference. Diagnostically, these patients appeared to fit the characterization of the group of pseudoneurotic schizophrenics described by Hoch and Polatin (1949).

2–7 Type ($N = 13$)

The following set of rules was developed to specify the 2–7 MMPI code type:

1. *D* and *Pt* over *T* score 70.
2. *D* greater than *Pt*.
3. *Pt* minus *Sc* at least 15 *T* scores.
4. *Si* below *T* score 70.
5. *Pd* and *Pa* below *T* score 80 unless *D* and/or *Pt* not over 85 in which instance *Pd* and *Pa* must be below 70.

Patients obtaining profiles of the 2–7 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: anxiety, anxiety reaction diagnosis, depression, insomnia, married, obsessions, tension, and worrying. Significantly different in frequency from patients-in-general at the .05 level were: cardiac complaint, loss of appetite, nervousness, rumination, and weak-tired-fatigued. Diagnostically these patients seemed to fit the description of the anxiety reactions or obsessive–compulsive personalities.

2–7–4 Type ($N = 27$)

The following rules were developed to specify the 2–7–4 type:

1. *D*, *Pd*, and *Pt* above *T* score 70.
2. *Pt* minus *Sc* at least 5 *T* scores.
3. *Pt* minus *D* 10 *T* scores or less.
4. *D* minus *Pd* 20 *T* scores or less.
5. Among scales *D*, *Pd*, and *Pt*, two scores below *T* score 100.
6. If *Pd* is peak, include only if *Pd* does not exceed *D* or *Pt* by more than 10 *T* scores.
7. *Si* below *T* score 70.
8. *Ma* less than *Sc*.
9. *Ma* above *T* score 40.
10. If *Ma* is below *T* score 50 include only if *Si* does not exceed *Ma* by more than 15 *T* scores.

Patients obtaining profiles of the 2–7–4 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: depressive reaction diagnosis, financial difficulty, heavy drinking, and tension. Significantly different from patients-in-general at the .05 level were: anxiety, conflict with wife, and insomnia. Diagnostically these patients appeared to fit the stereotype of the oral-dependent alcoholics.

Items which were less frequent for any profile type than for patients-in-general can also be found in Table 2. It will be noted that patients-in-general tend to be described more frequently as evasive and defensive than either the 2–7–8 or the 2–7–4 profile types.

Table 3 shows the check list items which differed significantly between the various combinations of profile types. These differences occur despite the fact that all three profiles are of the same two-point code type.

Discussion

The rules for specifying profiles have been presented in detail. With experience, deviations from the specified configurations are interpretable. For example, 2–7–8s with higher 4 (*Pd*) than permitted by the rules tended to have the general characteristics of the type but also tended to have histories of acting-out which frequently included police contacts and penitentiary sentences. With 6 (*Pa*) higher than permitted by the rules, an increase in hostility, in suspiciousness and, particularly, evidence of overt psychosis was found. In short, inferences which could be made from single scale or pair-wise scale deviations from the basic configuration could be used to modify the interpretation of the profile. Similarly, although age and intelligence were restricted in deriving the descriptions, cases which deviated on these variables tended to fit less well but usually were not complete misses. The rules have been devised to select cases which are highly purified for research purposes but in many instances the cutting point necessarily was somewhat arbitrary and resulted in the exclusion of cases which would have fit fairly well. For example, for the 2–7–8 profile type, the rule of *D* minus *Sc* of 15 *T* scores or less (Rule 2) was set to exclude depressive reactions even though some pseudoneurotic schizophrenics may be missed by rigid application of this rule.

It should be noted on the one hand that the three MMPI profile types reported would be expected to show many common traits relating to anxiety and depression since they were all of the 2–7 code type; on the other hand, there are many clear-cut differences despite the fact that they are all of the same two-point code type. For the sake of brevity more extensive clinical narrative descriptions which elaborate on the constellations of traits and contain inferences derived from them have been omitted. However, such descriptions have been developed for clinical use.

Previous research (Gilberstadt, 1952) utilizing the CC index of MMPI profile similarity devised by Hathaway and Meehl (Sutton, 1952; Williams, 1952) served as a pilot study for the present research. The study of the 2–7–8 profile type is essentially a cross-validation of earlier findings. In both studies the cases obtaining profiles of this type were highly similar to the syndrome of pseudoneurotic schizophrenia described by Hoch and Polatin (1949).

Further validation of the profile types by the Q sort technique is planned. This approach, when applied with descriptions derived from MMPI *Atlas* (Hathaway and Meehl, 1951) case histories, has already been shown by Duker (1958) to be fruitful.

Current research includes the most frequent MMPI code types and extends in scope to construct validation with other tests.

Summary

Contingencies for specifying three frequently occurring types of MMPI profiles of the *D-Pt* variety have been developed from a large pool of Veteran's Administration general hospital psychiatric cases. These include a *D-Pt-Sc* pseudoneurotic schizophrenic type, a *D-Pt* obsessive–compulsive type, and a *D-Pt-Pd* alcoholic type. Check list data from social histories and psychiatric discharge summaries are presented for each of the types. These profile types can be used as a source

of reference for profile interpretation or as a means of classification for research purposes.

REFERENCES

CANTOR, J. M. Syndromes found in a psychiatric population selected for certain MMPI code endings. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

CRONBACH, L. J., & GLESER, GOLDINE. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.

DUKER, JAN. The utility of the MMPI *Atlas* in the derivation of personality descriptions. Unpublished doctoral dissertation, Univer. Minnesota, 1958.

GILBERSTADT, H. An exploratory investigation of the Hathaway-Meehl method of MMPI profile analysis with psychiatric clinical data. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univer. Minnesota Press, 1951.

HOCH, P., & POLATIN, P. Pseudoneurotic forms of schizophrenia. *Psychiat. Quart.*, 1949, **23**, 248–276.

LAWSHECH, C. H., & BAKER, P. C. Three aides in the evaluation of the significance of the difference between two percentages. *Educ. psychol. Measmt.*, 1950, **10**, 263–270.

LYKKEN, D. T. A method of actuarial pattern analysis. *Psychol. Bull.*, 1956, **53**, 102–107.

MEEHL, P. E. *Clinical vs. statistical prediction*. Minneapolis: Univer. Minnesota Press, 1954.

MEEHL, P. E. Wanted—a good cookbook. *Amer. Psychologist*, 1956, **11**, 263–272.

MEEHL, P. E. A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *J. counsel. Psychol.*, 1959, **6**, 102–109.

MOSEL, J. N., & ROBERTS, JUNE. The comparability of measures of profile similarity: An empirical study. *J. counsel. Psychol.*, 1954, **18**, 61–66.

ROSEN, A. Development of some new MMPI scales for differentiation of psychiatric syndromes within an abnormal population. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

SUTTON, M. L. Profile patterning and descriptive correlates of patients having low scores on scale 9 of the MMPI. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

WILLIAMS, H. L. Differential effects of focal brain damage on the MMPI. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

(Received September 14, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 368–369

THE EFFECTIVENESS OF THE BENDER–GESTALT IN DIFFERENTIATING A FLIGHT GROUP FROM AN AGGRESSIVE GROUP OF ADOLESCENTS

LOREN V. COROTTO AND ROBERT H. CURNUTT
Napa State Hospital, Imola, California

There is some evidence to suggest that the Bender-Gestalt measures some aspect of ego strength. Pascal and Suttell (1951) developed an objective scoring system for the B-G that consists of tabulating certain deviations in the reproductions of the designs. These authors believe that "if . . . we limit our definition of ego functioning to the ability to reproduce faithfully the B-G drawings as presented, then we may say that ego strength lies on a continuum from very low to very high B-G scores" (p. 9). Therefore, the lower the score obtained on this quantitative continuum the greater is the individual's ego strength. One of the authors in another study was unable to relate $F+\%$ on the Rorschach with a Z score on the B-G (Curnutt & Lewis, 1954). In this study, however, an attempt will be made to follow up those results and, at least by inference, to see if it can be determined whether the B-G does measure some aspect of ego strength when used with behavioral criterion.

It is frequently noted in clinical practice that adolescents who tend to react aggressively to their environment often produce better B-G records than do adolescents who tend to utilize escape or flight behavior. It is generally assumed in the literature that people who utilize aggressive behavior tend to have more ego strength than individuals who tend to rely upon flight behavior. If the B-G does measure an aspect of ego strength, it should tend to differentiate between aggressive individuals and individuals who utilize flight. It is precisely this assumption which will be dealt with in this study.

The hypothesis to be tested may be stated as follows: The Pascal and Suttell scoring system of the B-G test will not significantly differentiate adolescents who react primarily with flight behavior in contrast to those who react primarily with aggressive behavior.

Procedure

The records of 92 adolescent patients who had been hospitalized in a children's psychiatric unit were withdrawn from the files. The patients in the aggressive group and in the flight group were matched individually on the basis of age, sex, and education. This was done since it is already known that both age and education are related to B-G scores (Bender, 1938; Pascal & Suttell, 1951). All individuals who had either a diagnosis or a known history of psychosis, mental deficiency, or organic brain damage were excluded.

The flight group was composed of the adolescents who had a record of two or more runaways from home since the age of 10 but with no resultant aggressive behavior. The aggressive group was defined as those subjects who had a record of two or more physically aggressive acts: e.g., destruction of property, fighting, or physical assault, but without any runaway behavior.

The selected sample was independently matched by one of the authors prior to obtaining the B-G protocols. The other author was subsequently presented with the randomized and codified protocols and asked to score them in the blind.

The sample utilized in the present study consisted of 27 matched pairs of males and 19 matched pairs of females. The ages of the sample ranged from 11 to 18. The Sign Test was used to analyze the obtained results.

Results and Discussion

The analysis of the 46 pairs of adolescents failed to yield a significant difference, thus sustaining the null hypothesis. However, when the paired individuals were examined on a sex basis, significant differences were obtained: i.e., the male scores had tended to cancel out the female scores. Hence the results can be explained purely on the basis of sex differences. In view of this finding, an

analysis restricted to the 27 individually matched pairs of males in the original sample was undertaken with the result that the B-G did significantly differentiate between the aggressive group and the flight group. The difference was significant at the .001 level when applying a two-tailed test of significance. In other words, the aggressive males had lower B-G scores, i.e., more ego strength as operationally defined by Pascal and Suttell, while the flight males had the higher B-G scores.

When the 19 individually matched pairs of females in the original sample were examined, an inverse relationship to that of the males was obtained. The difference was significant at the .001 level. In contrast to the male sample, the female flight group was found to have the lower B-G scores (more ego strength) while the aggressive female group obtained the higher B-G scores. The marked reversal of the behavioral correlates associated with the lower B-G scores by the female sample in comparison to the male sample would seem to emphasize caution in attempts to relate behavior to B-G scores without thoroughly allowing for sex differences. Since Pascal and Suttell (1951) were unable to demonstrate a significant sex difference in the standardization of their scoring system, the possibility is raised that social and cultural influences may affect the behavioral expression of that aspect of ego strength tapped by the B-G.

These findings may have implications for the treatment of adolescents whose behavioral symptoms primarily revolve around aggression and flight. Since the results indicate that boys who utilize flight reactions have relatively less ego strength than do aggressive boys, then possibly the therapist might wish to direct his efforts more towards supporting and developing the ego of the individual. In the aggressive group of boys, however, the aims of the therapist might be different. In this case, the therapist's efforts might be directed more towards channeling unacceptable behavior into more socially acceptable modes of response. Similar considerations would seem applicable to a female population.

Summary

In order to explore the effectiveness of the B-G in differentiating a group of adolescents who react primarily in an aggressive manner from a group which reacts primarily in a flight manner, 46 pairs of adolescents were matched by age, sex, and education. When the groups were segregated on a sex basis, significant differences (.001) were obtained. Adolescent girls who utilize primarily flight behavior tend to have lower B-G scores (more ego strength as operationally defined by Pascal & Suttell, 1951) than do adolescent girls who react primarily with aggressive behavior. Adolescent boys who tend primarily to react with aggressive behavior, however, tend to have lower B-G scores (more ego strength) than do adolescent boys whose primary behavior reactions are characterized by running away. Some implications of these findings are discussed.

REFERENCES

Bender, L. A. A visual motor Gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. res. Monogr.*, 1938. No. 3.

Curnutt, R. H., & Lewis, W. B. The relationship between *Z* Scores on Bender-Gestalt and *F* + % on the Rorschach. *J. clin. Psychol.*, 1954, **10**, 96–97.

Pascal, G. R., & Suttell, B. *The Bender-Gestalt test*. New York: Grune & Straton, 1951.

(Received August 4, 1959)

Journal of Consulting Psychology
1960, Vol. 24, No. 4, 370–374

BRIEF REPORTS

THE EPPS PROFILE STABILITY COEFFICIENT

DONALD W. FISKE, KENNETH HOWARD, AND WILLIAM RECHENBERG
University of Chicago

An unusual and almost uninvestigated feature of the Edwards Personal Preference Schedule (Edwards, 1957) is the Profile Stability Coefficient, the correlation (over needs) of the two sets of part scores for each *S*. This correlation was obtained for each of 135 college students. While Edwards uses *r* as the stability coefficient, we used rho, since the hierarchy of the needs seems of primary concern. Edwards' distribution and ours were highly congruent, even to a second mode at .85–.89. Our mean, obtained via a rho-squared transformation, was .73, while his was .74.

Inspection indicates that the stability coefficient is in part a reflection of the relative scatter or differentiation of the *S*'s profile of needs. It is also associated with separate scale scores. (In these analyses we used the square of rho to reduce the skewness of the distribution of stability indices). For EPPS variables, the stability score correlated .37 with n Autonomy, .30 with n Exhibition, .23 with the Consistency score, and .17 with both n Achievement and n Dominance; it also correlated −.28 with n Abasement, −.27 with n Order, and −.21 with n Deference. (.22 is significant at the .01 level, .17 at the .05).

Among 19 other test scores, it correlated .22 with Dominant and .18 with Active on the Thurstone Temperament Schedule. It also correlated .19 with Expressed Control and −.22 with Wanted Control scores from the FIRO-B Scale (Schutz, 1958, pp. 57–65). (No association was found with two tests of intelligence). These correlations are consistent with those for the EPPS scores. Consistency of a need profile on two groups of items appears to be related to active strivings for superior and controlling status.

For the same *S*s, the relationship between positive self-ratings and intra-individual variability was negative for each of six traits, although no dependable relationships were found between such variability and the same set of test scores (van der Veen & Fiske, 1960). However, the temporal variability of mean peer ratings on assertiveness and number of useful ideas contributed to a group problem solving session was negatively related to self-descriptions of assertiveness and sociability (Fiske, 1960). These several findings suggest that individuals who describe themselves in positive and active terms may be less variable in both their test responses and their behavior.

REFERENCES

EDWARDS, A. L. *Manual for The Edwards Personal Preference Schedule.* (Rev. ed.) New York: Psychological Corp., 1957.

FISKE, D. W. Variability among peer ratings in different situations. *Educ. psychol. Measmt.*, 1960, **20**, in press.

SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

VAN DER VEEN, F., & FISKE, D. W. Variability among self-ratings in different situations. *Educ. psychol. Measmt.*, 1960, **20**, 83–93.

(Received September 25, 1959)

MANIFEST NEEDS AND MANIFEST ANXIETY [1]

CARL D. WILLIAMS, ROBERT B. TALLARICO, AND JAMES T. TEDESCHI
University of Miami

The present study was designed to contribute additional empirical properties of "anxious" and "nonanxious" groups on the Taylor Manifest Anxiety Scale (MAS) by comparing their manifest needs as measured by the Edwards Personal Preference Schedule (EPPS). The MAS was given to 543 college students, and extreme groups of 65 persons each were identified. The EPPS was given to the 38 MAS high scorers and the 39 MAS low scorers who were willing to participate upon invitation. There were 23 male and 15 female high scorers and 19 male and 20 female low scorers. The MAS scores of the anxious group ranged from 21 to 34. The means for the participants and the nonparticipants from the 65 high MAS scorers were 25.92 and 26.78, respectively. This difference was not statistically significant. The MAS scores of the nonanxious group ranged from 0 to 4. The means for the participants and the nonparticipants from the 65 low MAS scorers were 2.46 and 2.31, respectively. This difference was not statistically significant.

Significant differences were found between high and low MAS scorers for three of the manifest needs in men and for five of the manifest needs in women. For men, the mean EPPS scores for the low and high groups, respectively were 8.74 and 11.78 for succorance ($p < .05$), 18.58 and 14.35 for dominance ($p < .01$), and 8.47 and 12.26 for abasement ($p < .01$). For women, similar comparisons were 13.55 and 10.87 for deference ($p < .05$), 19.20 and 16.13 for intraception ($p < .05$), 9.90 and 15.47 for abasement ($p < .01$), 12.85 and 8.20 for endurance ($p < .05$), and 10.25 and 13.27 for aggression ($p < .05$).

Abasement is the only manifest need for which significant differences exist for both men and women in the present study. The finding that both men and women nonanxious subjects scored significantly lower than anxious subjects on abasement suggests the possibility that the nonanxious groups may be characterized by a more favorable self-concept than the anxious groups.

The data for five of the six remaining EPPS variables on which significant differences were found were compared with Edwards' normative data. Edwards has reported significant differences between men and women in his normative data for all five of the variables compared. These comparisons showed that high MAS scorers differed from low MAS scorers on the variables of succorance and dominance for men and deference, intraception, and aggression for women in that the means of the anxious groups were more similar to the scores made by the opposite sex than to those made by the same sex in Edwards' normative group.

This result and the finding of lower scores on abasement for the nonanxious subjects appear to support the view that, as compared to low scorers, high scorers on the MAS are persons who are not as interested in making a good impression on others and may thus tend to say deviant things about themselves.

(Received September 28, 1959)

[1] This investigation was supported in part by a grant from the Penrose Fund of the American Philosophical Society. An extended report of this study may be obtained without charge from Carl D. Williams, Department of Psychology, University of Miami, Coral Gables, Florida, or for a fee from the American Documentation Institute. Order Document No. 6290 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

THE INFLUENCE OF ORDER OF ADMINISTRATION ON SELF–CONCEPT MEASURES [1]

DAVID S. PALERMO
Institute of Child Development and Welfare, University of Minnesota
AND JOHN G. MARTIRE
Southern Illinois University

Many studies have reported relationships between measures of actual- and ideal-self-concept and have frequently related measures of self-concept to other indices. This study is designed to investigate the effects of the order of administration of actual- and ideal-self-concept measures on their relationship to each other and their relationship to a third measure.

Self-concept measures were administered to 120 male and 120 female introductory psychology students at Southern Illinois University. Each *S* was given three trait ranking forms and a set of anchoring definitions for 26 traits used in previous studies by the second author. The *S*s were read the anchoring definitions and then instructed to read the directions at the top of the first sheet and make their rankings accordingly. When they finished the first set they were to go on to the second and then the third. Differential instructions at the top of the three ranking sheets directed *S*s to rank the traits so as to provide an assessment of self-ideal (sI), actual-self (As) and social-self (Ss). Half of the female and half of the male *S*s responded to the rankings in the order As, sI, and Ss (Order I), and the other half of the *S*s responded in the order sI, As, Ss (Order II). These orders were selected since most current interest is in ideal- and actual-self concepts.

Pearson *r*'s obtained between the two orders by sex group for the mean rankings of the 26 traits of the three scales ranged between .82 and .91 (all significant $< .01$ level). These correlations indicate that the relative position of each trait tended to remain the same on all three scales for both sexes regardless of order of administration.

Interscale correlations for each sex group under each order ranged between .26 and .88. These correlations reached the .05 level of confidence with but one exception. The male correlation under Order I between As and Ss (.26) was not significant. Of particular interest are the correlations between Ss and the other two measures under the two orders. Ss correlated .70 and .88 for males and females, respectively with the sI scale when the sI scale was given second, but only .58 and .74 when the sI scale was given first. The correlations between Ss and As were .52 and .68 for males and females, respectively when As was given second and .26 and .62 when As was given first. In each of the comparisons the correlation between the Ss scale, given third in all cases, and the other scales varied as a function of the order of administration. Although the difference between the two correlations was not significant in any of the four comparisons, the consistency of the relationship would suggest that an intervening scale tends to reduce the relationships obtained. If this trend is substantiated by additional research, doubt may be cast upon the validity of discrepancy scores as conventionally computed.

The correlations between the sexes on the rankings of the traits for the three scales ranged between .81 and .89 (all significant $< .01$ level). These data indicate that males and females show a great deal of similarity in their rankings of the traits on all three scales regardless of order of administration. The correlations obtained between sexes in Order II are strikingly similar to those previously obtained by Martire and Hornberger, although the interscale correlations for the two sexes were somewhat different than those obtained in the earlier study.

(Received October 5, 1959)

[1] An extended report of this study may be obtained without charge from David S. Palermo, Institute of Child Development and Welfare, University of Minnesota, Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. Order Document No. 6292 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

WECHSLER–BELLEVUE PATTERNS OF PSYCHOPATHS[1]

LEONARD KINGSLEY
Walter Reed Army Hospital, Washington, D. C.[2]

Wechsler (1944) reported that psychopaths generally obtain higher IQs on the Performance (P) Scale of his test (W-B) than on the Verbal (V) Scale and also described a pattern of subtest characteristics for psychopaths. There have been several W-B studies of psychopaths and delinquents with differing findings. Wechsler (1958) again offers the same IQ and subtest characteristics, qualifying his observations to adolescent sociopaths, but then adding "further experience has shown that it is also applicable to the adult male psychopath."

Twenty psychopathic prisoners in a military disciplinary barracks were matched individually for age and education with a group of prisoners considered to possess "no psychiatric disease." Both groups were later combined into a general prisoner category because of the similarity of the results and matched individually for education and intelligence with "normal" controls functioning in a military setting.[3]

P IQ was greater than V IQ, at 1% level of confidence, in all 3 groups. This difference was not significantly greater in the psychopaths as compared to the other prisoners nor in the general prisoner group as compared with the controls.

The pattern of subtests postulated by Wechsler was then considered to represent a theoretical rank order of subtest. The Average Spearman rank order correlation between the individual or combined Wechsler scores and the theoretical rank order was then computed for all 3 groups; the correlation was significant for all 3. The difference scores between each matched subject of each group for the subtests were then computed to see if this correlation was significantly greater in one group vs. the other; these differences were not significant.

Finally, the attempt was made to see if any of the individual W-B subtests, disregarding rank order, differentiated one group from the other; none of the subtests was significantly different in the psychopath vs. prisoner groups or in the general prisoner vs. control groups.

The results of this study raise doubts about the use of difference scores and subtest patterns on the W-B for psychopaths. Wechsler's observations as to the difference between the P and V IQs and the order of subtest difficulty were found to be present for the psychopaths, but these differences and patterns were also present for the nonpsychopathic prisoners and the "normal" controls. The same general trend existed in all three groups, with no significant differences. It is concluded that the IQ differences and subtest patterns offered by Wechsler and based upon the results of adolescent psychopaths cannot be extended to adult psychopaths in a military setting.

[1] An extended report of this study may be obtained without charge from the Secretary, Clinical Psychology Service, VA Hospital, Brooklyn 9, N. Y. or for a fee from the American Documentation Institute. Order Document No. 6294 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[2] Now at VA Hospital, Brooklyn, N. Y.

[3] The author is indebted to Ernest K. Montague for permission to use these W-B protocols.

REFERENCES

WECHSLER, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore: Williams & Wilkins, 1958.

(Received October 15, 1959)

THE FACTORIAL VALIDITY OF ITEMS ON THE IPAT ANXIETY SCALE[1]

A. W. BENDIG
University of Pittsburgh

The IPAT Anxiety Scale (Cattell, 1957) is composed of 40 items divided into five subscales with each subscale assumedly measuring one of Cattell's five oblique first-order factors (Q3, C, L, O, and Q4) whose intercorrelations define Cattell's second-order factor of "anxiety" (UI 24). These items were assembled into this new scale on the basis of their factor loadings in other inventories, and an analysis of the factor structure of the items in this form has not been reported to date.

The scale was anonymously administered to 200 undergraduate and graduate students in educational psychology classes (100 men and 100 women, mean age 27.8 years), and the items were intercorrelated by the usual product-moment method. Five factors were extracted by the complete centroid method, and the factors were rotated to oblique simple structure using the analytic criterion of rotation developed by Pizka and Saunders. The correlations among the oblique first order factors were factored and the second-order factors also rotated to oblique simple structure.[2]

The obtained item factor loadings showed little relationship to the assumed factor content, and none of Cattell's five factors could be clearly identified in this analysis. Of the 40 items 16 had no loadings above .29 on any of the factors. The correlations among the primary factors ranged from .33 to −.12, and two second-order factors were found instead of the unitary UI 24 factor expected with the correlation between these two oblique factors being approximately .01. Bifactor graphs indicated that further hand rotation would not materially improve simple structure or increase similarity to Cattell's assumed factor structure.

Internal consistency reliabilities were obtained for four of the scores suggested in the test manual (Cattell, 1957): Total Anxiety, Covert Anxiety, Overt Anxiety, and the Difference Score (Bendig, 1959). Using Kuder-Richard Formula 20 the reliabilities of these scores for this sample of 200 *Ss* were .81, .63, .76, and .24.

The results, unless attributable to a sampling or methodological artifact, strongly suggest that the IPAT Anxiety Scale does not measure what it purports to measure, namely the unitary UI 24 second-order factor of anxiety, but instead confounds two independent second-order factors.

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 6295 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[2] The author's appreciation is extended to Gary Lotto and William B. Kehl of the University's Computation and Data Processing Center for providing facilities for the statistical analysis.

REFERENCES

BENDIG, A. W. College norms for and concurrent validity of Cattell's IPAT Anxiety Scale. *Psychol. Newsltr, NYU*, 1959, **10**, 263–267.

CATTELL, R. B. *Handbook for the IPAT Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

(Received October 16, 1959)

ERRATUM

In our article, "Intra Q Deck Relationships as Influences and Realities in Personality Assessment," this JOURNAL, 1960, **24**, 61–66, Table 3, column 2, line 1 should be I** instead of II**. This error represented a reversal of the correct interpretation of the relationship between sizes of two correlations. In addition, the third sentence under the subheading, "Neurotic Stereotype and Personality test findings," p. 63, contains an interpretation of this reversal and should read . . . "was described as *less* neurotic . . ." rather than "more." The remaining materials in the paragraph should then be discounted. Robert D. Wirt and Peter F. Briggs.

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition the following special branches are dealt with fully: psychoanalysis, individual and analytical psychology; sexology, criminology, alcoholism, and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry, vocational guidance and mental testing; psychometrics; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1000 pp., including an index classified by author and subject

Subscription price $22.50 per annual volume

EXCERPTA MEDICA FOUNDATION

The New York Academy of Medicine Building

2 East 103 Street New York 29, N.Y.

THE TOTAL CHILD

AT DEVEREUX SCHOOLS a multidisciplined approach to evaluation, study, treatment, education, and rehabilitation focuses on understanding the total child and his needs. Instruction is highly individualized to take advantage of the specialized learning techniques most effective with (a) children of normal or superior intelligence who have emotional problems; (b) children with educational disabilities, due to mental retardation, brain injury, aphasia, dyslexia, and other similar conditions. The program extends from kindergarten through high school and beyond. Vocational training is also available.

CLINICAL STAFF

J. Clifford Scott, M.D.
Edwin E. Abrahamsen, M.D.
Aurelio Buonocore, M.D.
Charles H. Campbell, Jr., M.D.
Fred J. Calaman, M.D.
Ruth E. Duffy, M.D.
William F. Haines, M.D.
Robert L. Hunt, M.D.
Richard H. Lambert, M.D.
Leonardo Magrun, M.D.
Joseph J. Peters, M.D.
Alva J. Scott, M.D.
James S. Sherman, M.D.
Albert S. Yerdon, M.D.
Walter M. Uhler, M.D.
Theo L. Vlessis, M.D.

Lance Wright, M.D.
F. Ellsworth Henry, S.T.D.
Milton Bratten, Ph.D.
William J. Cohen, Ph.D.
Dorothy E. Conrad, Ph.D.
Sidney L. Copel, Ed.D.
Shirley M. Johnson, Ph.D.
Noel Jenkin, Ph.D.
John E. Kleiner, Ph.D.
Murray Levine, Ph.D.
Henry Platt, Ph.D.
Edgar A. Smith, Ed.D.
George Spivack, Ph.D.
Herbert A. Spright, Ph.D.
Anne Howe, M.A.
Kenneth E. Dunn, B.S.

Psychoanalytic Consultants

G. Henry Katz, M. D.
Herbert H. Hornbostle, M.D.

THE DEVEREUX FOUNDATION

A nonprofit organization — *Founded 1912*

Devon, Pennsylvania

Santa Barbara, California — *Victoria, Texas*

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX
Administrative Consultant

EDWARD L. FRENCH, Ph.D.
Director

WILLIAM B. LOEB
Treasurer

Professional inquiries for Eastern Schools should be directed to Charles J. Fowler, *Director of Admissions*, Devereux Schools, Devon, Pa.; for Pacific Coast Schools, to Keith A. Seaton, *Registrar*, Devereux Schools in California, Santa Barbara, Calif.; Southwestern residents address Devereux Schools in Texas, Box 336, Victoria, Tex.